NASA Technical Memorandum 104566, Vol. 20

SeaWiFS Technical Report Series

Stanford B. Hooker and Elaine R. Firestone, Editors

Volume 20, The SeaWiFS Bio-Optical Archive and Storage System (SeaBASS), Part 1

Stanford B. Hooker, Charles R. McClain, James K. Firestone,
Todd L. Westphal, Eueng-nan Yeh, and Yuntao Ge

JULY 1994

NASA Technical Memorandum 104566, Vol. 20

SeaWiFS Technical Report Series

Stanford B. Hooker, Editor
NASA Goddard Space Flight Center
Greenbelt, Maryland

Elaine R. Firestone, Technical Editor
General Sciences Corporation
Laurel, Maryland

Volume 20, The SeaWiFS Bio-Optical Archive and Storage System (SeaBASS), Part 1

Stanford B. Hooker
Charles R. McClain
NASA Goddard Space Flight Center
Greenbelt, Maryland

James K. Firestone
Todd L. Westphal
Eueng-nan Yeh
General Sciences Corporation
Laurel, Maryland

Yuntao Ge
University Space Research Association
Columbia, Maryland

PREFACE

The Sea-viewing Wide Field-of-view Sensor (SeaWiFS) is the successor ocean color imaging system to the NIMBUS-7 Coastal Zone Color Scanner (CZCS). The SeaWiFS calibration and validation effort (McClain et al. 1992) includes laboratory, on-orbit, *in situ*, and quality control elements which require pre- and post-launch measurements and analyses. The CZCS experience demonstrated the need for comprehensive documentation and access to digital data from the prelaunch sensor calibration and characterization, high quality *in situ* data for bio-optical algorithm development (Clark 1981, Mitchell and Holm-Hansen 1991), and post-launch vicarious calibration (Evans and Gordon 1994). The SeaWiFS calibration and validation program is designed to support the collection and archiving of these data.

The laboratory element is concerned with several activities in support of producing verifiable radiances from the instrument. These activities include the prelaunch characterization of the sensor [Barnes et al. (1994)] and a continuing series of SeaWiFS Intercalibration Round-Robin Experiments (SIRREXs) held once a year (Mueller 1993 and Mueller et al. 1994). The prelaunch laboratory effort involves establishing and updating calibration procedures for SeaWiFS data, bio-optical and atmospheric algorithm development, and quality control of all of the relevant data sets. The SIRREX program involves the transfer of National Institute of Standards and Technology (NIST) scales of spectral irradiance and radiance, through the SeaWiFS Project, to all participating laboratories in the science community and to the calibration standards used to calibrate the SeaWiFS instrument for radiance responsivity. Data from the SIRREX and prelaunch SeaWiFS calibration provided by the Hughes/Santa Barbara Research Center (SBRC) and Orbital Sciences Corporation (OSC) are included in the calibration and validation database.

The on-orbit sensor calibration is composed of several types of measurements including solar, lunar, and internal electronic data collection (Woodward et al. 1993). In addition, the sensor engineering data will be routinely monitored. These data and related analysis products will be archived within the calibration and validation database in order to track long term system performance.

The *in situ* element incorporates measurements of water-leaving radiances and other related quantities, from ships and moorings, to develop a vicarious calibration data set based on geographically diverse samples of oceanic and atmospheric data sets. The bio-optical database includes historical pigment and optical data, e.g., CZCS NIMBUS Experiment Team (NET) data, which will be expanded over time from 1) deployments of the project-sponsored Marine Optical Characterization Experiment (MOCE) Team, 2) an optical time series collected from a marine optical buoy (MOBY), and 3) other field expeditions producing data in keeping with an evolving set of optical protocols (Mueller and Austin 1992).

For these measurements to be of use to the pre- and post-launch activities associated with calibrating and validating SeaWiFS data, a well documented archive which is easily accessed by all potential users is needed. This requirement has produced the SeaWiFS Bio-Optical Archive and Storage System (SeaBASS); this volume of the SeaWiFS Technical Report series is the first step in documenting the architecture of this database.

Greenbelt, Maryland
May 1994

— C. R. McClain
and S. B. Hooker

Table of Contents

ABSTRACT

This document provides an overview of the Sea-viewing Wide Field-of-view Sensor (SeaWiFS) Bio-Optical Archive and Storage System (SeaBASS), which will serve as a repository for numerous data sets of interest to the SeaWiFS Science Team and other approved investigators in the oceanographic community. The data collected will be those data sets suitable for the development and evaluation of bio-optical algorithms which include results from SeaWiFS Intercalibration Round-Robin Experiments (SIRREXs), prelaunch characterization of the SeaWiFS instrument by its manufacturer—Hughes/Santa Barbara Research Center (SBRC), Marine Optical Characterization Experiment (MOCE) cruises, Marine Optical Buoy (MOBY) deployments and refurbishments, and field studies of other scientists outside of NASA. The primary goal of the data system is to provide a simple mechanism for querying the available archive and requesting specific items, while assuring that the data is made available only to authorized users. The design, construction, and maintenance of SeaBASS is the responsibility of the SeaWiFS Calibration and Validation Team (CVT). This report is concerned with documenting the execution of this task by the CVT and consists of a series of chapters detailing the various data sets involved. The topics presented are as follows: 1) overview of the SeaBASS file architecture, 2) the bio-optical data system, 3) the historical pigment database, 4) the SIRREX database, and 5) the SBRC database.

PROLOGUE

The Sea-viewing Wide Field-of-view Sensor (SeaWiFS) ocean color radiometer is scheduled to be launched in late 1994 with an expected lifetime of five years. The National Aeronautics and Space Administration's (NASA) Goddard Space Flight Center (GSFC) is tasked with carrying out a program to acquire the global SeaWiFS data set, validate and monitor its accuracy and quality, process the radiometric data into geophysical units using a set of atmosphereic and bio-optical algorithms, and distribute the final products to the scientific community through the GSFC Distributed Active Archive Center (DAAC). The SeaWiFS data products figure as prominent components of major scientific programs studying global change, including the Joint Global Ocean Flux Study (JGOFS), the World Ocean Circulation Experiment (WOCE), and the Global Ocean Ecosystems dynamics (GLOBEC) programs.

Two important goals of the SeaWiFS Project are to determine, from the satellite radiance measurements, a) normalized water-leaving radiance with an accuracy of 5%, and b) chlorophyll a concentration with an accuracy of 30% (Hooker et al. 1992). The amount of historical data suitable for developing and validating the algorithms employed is very limited because of the specific suite of simultaneous observations and the radiometric accuracies required. (Because the desired result is to validate spaceborne radiances to within 5%, the comparative *in situ* radiometric measurements must be calibrated to be accurate within less than 5%.)

For the Coastal Zone Color Scanner (CZCS), which collected data from 1978–86, few algorithms have been published and these rarely show broad agreement at the hoped for level. The algorithmic differences are primarily the result of two factors: 1) the data used are collected from regimes having different bio-optical properties and 2) the instrumentation and sampling techniques employed are different.

In order to minimize observational errors, the SeaWiFS Project sponsored a multidisciplinary workshop to outline the observations and sampling protocols required for bio-optical algorithm development (Mueller and Austin 1992). One of the consequences of the workshop was the establishment of a series of SeaWiFS Intercalibration Round-Robin Experiment (SIRREX) activities to establish and advance the state of the art for making validation measurements to verify the radiometric precision and long-term stability of the SeaWiFS instrument's radiance responsivities, and to validate the atmospheric correction models and algorithms used to convert SeaWiFS radiances to normalized water-leaving radiances. One of the principal approaches to this critical aspect of validation will be frequent direct comparisons between SeaWiFS estimates and *in situ* measurements collected during Marine Optical Characterization Experiment (MOCE) deployments and from the Marine Optical Buoy (MOBY).

The only economically feasible approach for minimizing spatial biases is to maximize the acquisition of a globally distributed sampling of *in situ* measurements by soliciting contributions of data from the oceanographic community at large, and to combine them with data collected from SeaWiFS sponsored field programs (MOCE and MOBY deployments) and instrument calibrations (SIRREX activities) in an easily accessed database. The database can then be used to formulate and compare advanced CZCS-type algorithms with the standard algorithms and to develop preliminary SeaWiFS chlorophyll-like pigment algorithms.

The acquisition of the historical and contemporary data sets requires the implementation of quality control (QC),

documentation, and cataloging procedures and the design of a database structure suitable for bio-optical data with a user friendly interface. This task is the responsibility of the SeaWiFS Calibration and Validation Team (CVT). This report is concerned with documenting the execution of this task by the CVT and consists of a series of chapters detailing the various data sets included in the SeaWiFS Bio-Optical Archive and Storage System (SeaBASS). A short synopsis of each chapter is given below.

1. Overview of the SeaBASS File Architecture

The SeaBASS file architecture is concerned with maintaining easy data access while accomodating a variety of operating systems. A text format is used with a simple file structure composed of two parts: a header block and a data block. The former is composed of a unique vocabulary expressed as a set of keywords and values; the latter is composed of a set of observables expressed as a tab-delimited matrix—the so-called spreadsheet format. A condensed file naming convention is used for systems utilizing a maximum of 11 characters; an expanded file naming convention, which is more readily deciphered, is also available when requesting data. To ensure a complete description of the data, additional format conventions are recommended for all data collection efforts.

2. The Bio-Optical Data System

The CVT has developed a software system which allows approved investigators to query and order holdings from SeaBASS. As of this writing, the archive includes data provided by SIRREX exercises, NIMBUS Experiment Team (NET) deployments, the Bermuda Atlantic Time Series (BATS), the Bermuda Bio-Optical Profiler (BBOP), the Center for Hydro-Optics and Remote Sensing (CHORS), WOCE/JGOFS activities, Hughes/Santa Barbara Research Center (SBRC) prelaunch calibration data, and over 50 historical pigment data sets. Data from MOCE cruises and MOBY will be added to the archive after SeaWiFS is launched. The CVT has made each of the data files in the archive self-describing by adding a metadata header which includes information such as the file's spatial and temporal limits, parameters, and investigators. The metadata is stored in a Sybase relational database, which also tracks information relating to system users and the status of data

orders. Authorized users will be able to query the contents of the archive through a graphical user interface (GUI) built with the Interactive Data Language (IDL) software and interfaced to the Sybase tables. This chapter describes the system file naming conventions and header keywords, database schema, data holdings, and user interface.

3. The Historical Pigment Database

The SeaWiFS calibration and validation historical pigment database was assembled for evaluating bio-optical algorithms, and the quality control of derived products from SeaWiFS and other satellite ocean color sensors. As of this writing, there are more than 330,000 independent pigment observations in the database, representing over 50 data sets amassed between 1956 and 1994 from a wide variety of investigators and oceanographic expeditions. This chapter describes the structure of the database, the current set of holdings, the procedures for submitting pigment data to the SeaWiFS Project, and the methods for converting data contributions from the data format in which they are submitted to the database format.

4. The SIRREX Database

This chapter describes the procedures used in obtaining and archiving the SIRREX data sets. These data sets provide a calibration history for the various lamps, spheres, plaques, and radiometers used as a radiometric standard for field measurements. Currently, there are two SIRREX data sets in the archive. This chapter contains the file naming and header protocol used for SIRREX data, the procedure for archiving the original data, and an example of an archived file.

5. The SBRC Database

This chapter describes the naming convention and procedures used in archiving the SeaWiFS prelaunch calibration data obtained from SBRC. The archive contains data from bright target recovery (BTR), absolute spectral response (ASR), line spread function (LSF), and temperature dependence measurements as well as images of the sun, moon, sky, and an integrating sphere. This data set contains characterization information on the SeaWiFS instrument provided by the manufacturer. As such, it serves as a baseline to the operation of the instrument on orbit.

Chapter 1

Overview of the SeaBASS File Architecture

STANFORD B. HOOKER
NASA Goddard Space Flight Center, Greenbelt, Maryland

JAMES K. FIRESTONE
General Sciences Corporation, Laurel, Maryland

ABSTRACT

The SeaBASS file architecture is concerned with maintaining easy data access while accomodating a variety of operating systems. A text format is used with a simple file structure composed of two parts: a header block and a data block. The former is composed of a unique vocabulary expressed as a set of keywords and values; the latter is composed of a set of observables expressed as a tab-delimited matrix—the so-called spreadsheet format. A condensed file naming convention is used for systems utilizing a maximum of 11 characters; an expanded file naming convention, using up to 31 characters which is more readily deciphered, is also available when requesting data. To ensure a complete description of the data, additional format conventions are recommended for all data collection efforts.

1.1 INTRODUCTION

To maintain flexibility and cross-platform access, the SeaBASS file architecture utilizes a condensed and expanded file naming convention. The former makes use of the so-called *eight-point-three* 11-character format designed for personal computers (PCs) running the Microsoft disk operating system (MS-DOS), which has the most restrictive naming convention commonly in use; the latter is for any computer supporting 31-character file names.

The condensed file naming convention considers the first eight characters as a prefix and the last three as a suffix:

PPPPPPPP.SSS

(where PPPPPPPP is the prefix and SSS is the suffix). The prefix and suffix are used to describe who collected the data, the type of data collected, and the circumstances surrounding data collection, i.e., when, where, etc.

All files in the database are stored in the condensed format, but users can request data in the expanded format, which is more readily deciphered. The expanded file name utilizes multiple prefixes to enhance readability and is a *verbose* version of the condensed file name.

Although upper- and lower-case letters are used in presenting the file conventions associated with the database, the actual codes and identifiers are not case sensitive, thus, a in a file name or header is the same as A. Most computer systems interpret non-alphanumeric characters as special characters, so allowed codes and identifiers used to build up a file name are restricted to numbers and letters, that is, $0, 1, \ldots, 9, A, B, \ldots, Z$ (hereafter referred to as $0, \ldots, Z$), as well as the underscore (_) and the period (.). When used for most sequential codes, e.g., the day of the month, the 0 entry is not used, so the allowed entries are $1, 2, \ldots, 9, A, B, \ldots, Z$ (hereafter referred to as $1, \ldots, Z$). In the expanded file naming convention, the period is used more frequently to enhance readability.

The basic file structure (Fig. 1) is composed of two parts: a header block and a data block. The former is composed of an agreed upon vocabulary expressed as a set of keywords with each keyword and its argument occupying one line in the header; the latter is composed of a set of observables expressed as a matrix with each row in the matrix individually numbered and terminated by a carriage return, and each data entry in the row separated by a tab—the so-called spreadsheet format. The header block begins with the /begin_header keyword and ends with the /end_header@ keyword.

```
/begin_header⟨cr⟩
  ⋮
/end_header@⟨cr⟩
1⟨tab⟩Value(1,1)⟨tab⟩Value(1,2)...Value(1,N)⟨cr⟩
2⟨tab⟩Value(2,1)⟨tab⟩Value(2,2)...Value(2,N)⟨cr⟩
3⟨tab⟩Value(3,1)⟨tab⟩Value(3,2)...Value(3,N)⟨cr⟩
  ⋮
M⟨tab⟩Value(M,1)⟨tab⟩Value(M,2)...Value(M,N)⟨cr⟩
```

Fig. 1. Basic database file structure.

1.2 ARCHIVAL PROCEDURES

The database protocol specifies a header format which is inserted at the beginning of every archived file. Any relevant information in an existing header is incorporated into the keyword vocabulary; additional notations are incorporated as comment lines, which begin with an exclamation point (!). The requirement for tab-delimited data and a sequence column often requires modification of the original data. While the data is being modified, a consistent numeric format is used for each data column, which is based on the range and precision of the data. Using a consistent numeric format allows a quick visual inspection of the data to locate errors.

Investigators submitting data to the database are encouraged to provide data in a form as close to the established protocol as is feasible. The use of commercial off-the-shelf (COTS) packages,† like Lotus 1-2-3 or Excel, is permitted as long as the basic tab-delimited American Standard Code for Information Interchange (ASCII) format is adhered to.

1.3 KEYWORDS

The keywords are divided into two groups: those required for every file and those approved for completing the description of the data in the file. One of the advantages of the keyword approach is once the vocabulary of the keywords is adopted and made unique, automated processing is enhanced because the files can be easily parsed. Locally defined keywords can be created, perhaps to track post-processing of the data.

A keyword line is made up of a keyword and, where required, its value:

$$\text{/keyword}=value\langle cr\rangle$$

where keyword is an approved keyword that must begin with a slash (/) and end with an equals (=) sign except for the keywords that denote the beginning and ending of the header block (see below); value is a string or numeric sequence giving the value to the keyword; and $\langle cr\rangle$ is a terminating carriage return that all line (or data row) entries must have. The value entries are represented by one of the following:

$$\langle n\rangle, \quad \langle s\rangle, \quad \langle n\rangle[u], \quad \text{or} \quad \langle s\rangle[u],$$

where $\langle n\rangle$ is a numeric entry composed of characters used to represent numbers, $\langle s\rangle$ is a character string, and $[u]$ is a string that gives the units of the keyword value.

In general, the International System of Units (SI), or *Système International d' Unitès*, are used except where

traditional usage dictates otherwise. For indicating numeric ranges, the right bracket (>) is used in place of the hyphen (-), which is reserved for negation, and the semicolon is used for the numeric increment. Thus, to indicate a measurement from 300–900 nm in steps of 10 nm, the entry would be 300>900;10[nm]. If the increment is one, the entire increment field can be left out. When constructing lists, commas are used to separate entries, and ranges are used where needed. To indicate all channels out of 6 except channel 2, for example, 1,3>6 is used. When denoting numbers raised to powers, the caret (^) is used with the exponent enclosed in braces, thus, a 10^{-2} entry is written as 10^{-2}. In writing units, the negative exponent is preferred; the solidus may be used to emphasize certain common unit combinations.

1.4 FORMAT CONVENTIONS

The following format conventions are recommended for all data collection efforts:

1. All files are stored and retrieved in the ASCII text format.

2. All files are self-describing and are composed of a header block followed by the data.

3. The data is in a tab-delimited spreadsheet format, that is, rows and columns with all column entries separated by a tab.

4. The header block is constructed from a unique vocabulary set and is made up of a series of one-line entries, each composed of a single keyword entry with its value, which is terminated with a carriage return. A listing of the allowed keyword entries is given in the keyword glossary below.

5. The first header block line contains the keyword /begin_header and the last header block line contains the keyword /end_header@. All of the intervening entries are keyword lines or comment lines and do not contain the backslash (\) or at-sign (@) characters.

6. The header block can contain comment lines. All comment lines start with an exclamation symbol (!) and, like keyword entries, do not contain the backslash (\) or at-sign (@), although they may contain white space (blanks).

7. All header information that is not part of a comment line is underscore filled, that is, underscores are used instead of blanks. This allows for easy parsing, since many text translation routines (inside and outside of COTS programs) replace tabs with blanks.

8. All entries are time stamped using the format used in the date and time keywords.

9. All entries start with a sequential sample number, so missing data or file corruption can be readily discerned.

---

† Identification of commercial software or equipment to specify adequately the experimental problem, does not imply recommendation or endorsement by the SeaWiFS Project, GSFC, or NASA, nor does it imply that the equipment identified is necessarily the best available for the purpose.

10. Any other information not given in the keyword entries that helps interpret the data, either by itself or in conjunction with other likely data sets, is included in the header as comment lines (see item 6 above) or as separate documentation files.

Chapter 2

The Bio-Optical Data System

JAMES K. FIRESTONE
EUENG-NAN YEH
TODD L. WESTPHAL
General Sciences Corporation, Laurel, Maryland

CHARLES R. MCCLAIN
STANFORD B. HOOKER
NASA Goddard Space Flight Center, Greenbelt, Maryland

YUNTAO GE
University Space Research Association, Columbia, Maryland

ABSTRACT

The CVT has developed a software system which allows approved investigators to query and order holdings from SeaBASS. As of this writing, the archive includes data provided by SIRREX exercises, NET deployments, BATS collection efforts, BBOP, CHORS, WOCE/JGOFS activities, SBRC prelaunch calibration data, and over 50 historical pigment data sets. Data from MOCE cruises and MOBY will be added to the archive after SeaWiFS is launched. The CVT has made each of the data files in the archive self-describing by adding a metadata header which includes information such as the file's spatial and temporal limits, parameters, and investigators. The metadata is stored in a Sybase relational database, which also tracks information relating to system users and the status of data orders. Authorized users will be able to query the contents of the archive through a GUI built with IDL software and interfaced to the Sybase tables. This document describes the system file naming conventions and header keywords, database schema, data holdings, and user interface.

2.1 INTRODUCTION

Figure 2 details the overall flow of data through the data system. The boxes along the top indicate each of the major classifications of data to be stored, including data from instrument calibration activities (SIRREX exercises and SBRC prelaunch measurements) and the *in situ* sampling program (MOCE and MOBY deployments).

After reception by the Project, each data set that is not SIRREX data is assigned a priority based on its (perceived) demand by the user community. Initially, a two-class priority scheme is envisioned. Data sets within the high-priority category, category A, will be subdivided, so that one file is created for each parameter or granule contained within the submitted original. This approach has several advantages:

1. It enables the data users to perform queries at the parameter level.
2. It ensures a unique file name for each granule of data.
3. It potentially reduces the size of the files the user will need to manipulate.

4. It simplifies the process of archiving data updates from the individual investigators.

Each category A data set will be visually checked to determine the optimum method for subdividing its contents. A program will then be written which performs the subdivision, fills missing data with a uniform code of -9999, converts time and space coordinates to use standard representations, and tab delimits the data columns. Data sets within the low-priority category (category B) will not be subdivided initially, and will retain their original format. At a later date, a category B data set may be converted to category A and subdivided as demand for it increases.

All output files will be named according to the naming conventions applicable to the data, as described in Section 2.4. The name includes the specification of a parameter, which will refer to a specific field for a category A data set or to a broader classification, e.g., experiment, for a category B data set. The current set of bio-optical parameters identified for the files that are not comprised of SIRREX data is given in Section 2.2.3.

After the files are named, a set of keywords are identified which completely describe the data set. The keywords
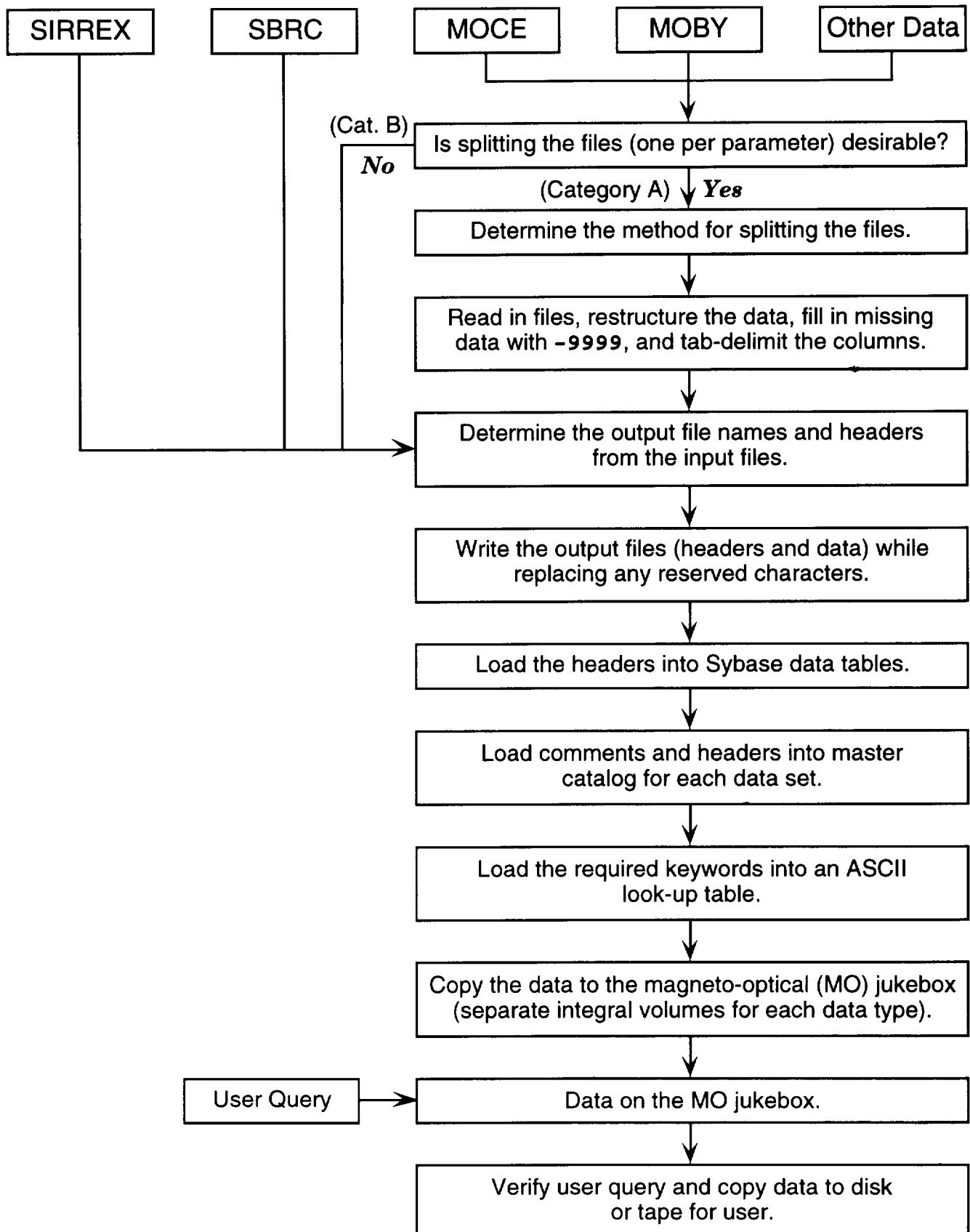
Fig. 2. The overall flow of data through the data system.

consist of both a required set, containing spatial and temporal information, column headings, and investigator(s), plus a more open-ended supplementary set approved for the data set. Section 2.5 lists possible values for both the required and supplementary sets. The header information is inserted at the beginning of each file. Following the header are any comments from the provider's original file(s), which may include the original headers. All comment lines are preceded in the output files by an exclamation character (!). The final step in the file creation is to remove all special characters, such as backslashes and boolean operators, which could be misinterpreted by the end user's computer operating system.

In order to facilitate the query of the metadata contained in the headers, their contents are loaded in three different places: 1) a set of Sybase relational database tables, 2) an ASCII *look-up* table containing just the required keywords in a columnar one line entry per file, and 3) a master catalog containing the headers and any internal comments for all files belonging to each unique data type. The current database table structure, which is designed to be of third normal relational form, is shown in Table 1. Note the set of required keywords for each major class of data (SIRREX, SBRC, or other bio-optical) are stored in separate tables, which serve as *look-up* tables in the database. Within these tables are foreign keys pointing to primary keys in other tables which store information on experiments, investigators, file names, etc.

There is also a set of administrative tables which track the users of the system, their requests for data, their access to the system, their comments on the system, and the Project's verification of their individual data orders (the fourth column in Table 1).

2.2 CONDENSED CONVENTION

The condensed file name follows the previously mentioned *eight-point-three* format and is of the form:

<div align="center">DDDPPPII.TNN</div>

(composed of a prefix, DDDPPPII, and a suffix, TNN). The prefix is composed of three parts: DDD is the date code, PPP is the position code, and II is the data identifier code. The suffix is composed of two parts: T is the type code and NN is the sequence code.

2.2.1 Condensed Date Code

The condensed date code for the data, DDD, indicates the number of days since 1 January 1950. Each column in the DDD entry consists of an alphanumeric entry incrementing from 0, . . . , 9, then A, . . . , Z. A code of 000 is reserved for the case where the date is unknown. A code of 001 represents 1 January 1950, 009 represents 9 January 1950, 00A represents 10 January 1950, 00Z represents 4 February 1950, 010 represents 5 February 1950, etc.

2.2.2 Condensed Position Code

The condensed geolocation position for the data, PPP, is based on a series of global 2.5° latitude by 2.5° longitude bins. Each column in the PPP entry consists of a letter in the range A, . . . , Z. The first bin, AAA, has its northwest corner at (−180,90); the second bin, AAB, has its northwest corner at (−177.5,90); the 145th bin, AFO, has its northwest corner at (−180,87.5). An entry of ZZZ is reserved for the case where the location is unknown.

2.2.3 Condensed Identifier Code

Each character in the condensed data identifier code, II, consists of either alphabetic (A, . . . , Z) or numeric values (0, . . . , 9). For category A data sets, each parameter or set of related parameters is saved as a separate file, and II is expressed as a value in the range AA, . . . , ZZ. Three of the characters, A, S, and Z, if present as the left-most character in the II code, are reserved for representing airborne, surface, and subsurface parameters, respectively. A value of AF, for example, represents airborne fluorescence data.

For category B data sets, II is the encoded name of the experiment or individual responsible for the collection of the data. For example, ON represents NET optical data. Table 2 contains a list of the experiments of the current data holdings, including the experiments referenced in the number II entries.

The condensed identifier codes with the corresponding verbose descriptions are as follows:

1M MOCE-I cruise data
2M MOCE-II cruise data
OW Marra WOCE/JGOFS data
OB Siegel BBOP data
1B Michaels BATS pigment data
OC CHORS optical data
OJ JGOFS EqPac data
ON NET optical data
1N NET pigment data

AF Airborne fluorescence
AH Air relative humidity
AI Airborne spectral downwelled irradiance
AP Atmospheric pressure
AR Airborne spectral upwelled radiance
AT Air temperature

BA Spectral beam attenuation coefficient
BC Beam attenuation coefficient
BS Spectral backscattering coefficient

CA Satellite pigment (chlorophyll *a*) concentration (computed)
CC Coccolith concentration
CD Diffuse attenuation coefficient, downwelled irradiance
CF Chlorophyll *a* fluorescence profiles
CP Chlorophyll *a* profiles

Table 1. Bio-optical database schema for storing: a) metadata common to all bio-optical data sets, and specific metadata for non-SIRREX data sets (left column); b) metadata specific to SIRREX data files (left middle column); c) metadata specific to SBRC data files (right middle column); and d) administrative information on system users (right column).

BIOOPT_LOOKUP
- Start date
- Start time
- End date
- End time
- Start latitude
- Start longitude
- End latitude
- End longitude
- [F] Investigator ID
- [F] Affiliation ID
- [F] Experiment ID
- [F] Data type ID
- Column headings
- Sequence number
- [F] Data file ID
- [F] Archive file ID
- [F] Original file ID
- [F] Detail record ID

INVESTIGATORS
- [P] Investigator ID
- [F] Affiliation ID
- Name
- Mailing address
- Telephone number
- E-mail address

DATA_TYPES
- [P] Data type ID
- Name

BIOOPT
- [P] Detail record ID
- Start time units
- End time units
- Start latitude units
- End latitude units
- Start longitude units
- End longitude units
- Documentation file

AFFILIATIONS
- [P] Affiliation ID
- Name

EXPERIMENTS
- [P] Experiment ID
- Name

DATA_FILES
- [P] Data file ID
- Name

ORIGINAL_FILES
- [P] Original file ID
- Name

ARCHIVE_FILES
- [P] Archive file ID
- Name

SIRREX_LOOKUP
- Column Headings
- [F] Configuration ID
- [F] Data type ID
- Date
- [F] Experiment ID
- [F] Testing group ID
- [F] Lamp/sphere/plaque ID
- [F] Instrument ID
- [F] Investigators ID
- Number of data rows
- [F] Original file ID
- [F] Run type ID
- [F] Archive File ID
- Sequence number
- [F] Detail record ID

SIRREX
- [P] Detail record ID (72 additional fields and units where applicable)

CONFIGURATIONS
- [P] Configuration ID
- Name

GROUPS
- [P] Testing group ID
- Name

LAMPS
- [P] Lamp/sphere/plaque ID
- Name

INSTRUMENTS
- [P] Instrument ID
- Name

RUN_TYPES
- [P] Run types ID
- Name

SBRC_LOOKUP
- Column Headings
- [F] Archive file ID
- [F] Data type ID
- Date
- [F] Instrument ID
- [F] Investigator ID
- Number of data rows
- [F] Original file ID
- [F] Run type ID
- Sequence number
- Band
- [F] Detail record ID

SBRC
- [P] Detail record ID
- Collection aperture
- Collection aperture units
- IFOV†
- IFOV units
- Counts per amp
- Counts per amp units
- Factor
- Filter
- Filter units
- Information file
- Lamp
- Modification status
- Slit
- Slit units
- Time
- Time units
- Mirror side
- Comment
- Wavelength range
- Wavelength range units

REQUESTS
- [P] Request ID
- Date and time of request
- [F] Local account ID
- [F] Archive file ID
- Date verification sent
- Date response received
- Date and time filed and copied from MO to disk or tape
- Date and time file sent to or copied by user

ACCOUNTS
- [P] Local account ID
- Account Name
- Long account name
- Mailing address
- Telephone number
- E-mail address
- [F] Preferred media type ID
- [F] Preferred verification type ID

LOGINS
- [P] Login ID
- [F] Local account ID
- Login date and time
- Logout date and time
- [F] Request ID
- [F] Comment ID

MEDIA
- [P] Preferred media type ID
- Name

COMMENTS
- [P] Comment ID
- [F] Local account ID
- Text of comment
- Date and time of comment
- Resolution, if a problem
- Date of resolution
- Person resolving problem

VERIFICATION
- [P] Preferred verification type ID
- Name

[P] = Primary key. [F] = Foreign key. † Instantaneous Field-of-View

Table 2. Currently held data sets (as of June 1994).

Date(s)	Name	Provider(s)	Location	Description
1992	MOCE-I§	Clark, Hooker, and Trees	Monterey Bay	Optics and Pigments
1993	MOCE-II§	Clark, Hooker, and Trees	Gulf of California	Optics and Pigments
1975–79	NET Stations	Clark	Coastal US	Optics and Pigments
1978–81	NET Along-Track	Clark	Coastal US	Optics and Pigments
1991	WOCE/JGOFS	Marra	Tahiti→Hawaii	Optics and Pigments
1992	JGOFS EqPac	Numerous	Equatorial Pacific	Optics and Pigments
1992–3	BBOP	Siegel	Bermuda	Optics
1991	CHORS/BOFS†	Trees	North Atlantic	Optics
1993–3	BATS	Michaels	Bermuda	Pigments
1992	SIRREX-1	Mueller, Hooker, and Johnson	CHORS	Instrument Calibration
1993	SIRREX-2	Mueller, Hooker, and Johnson	CHORS	Instrument Calibration
1994	SIRREX-3	Mueller, Hooker, and Johnson	CHORS	Instrument Calibration
1994	SBRC	Roberts and Holmes	SBRC	SeaWiFS Calibration
1956–94	Historical Pigments	Numerous	Global	>330,000 Observations

§Data sets will be made available after SeaWiFS launch.
†British Ocean Flux Study

CS Spectral absorption coefficient
CU Diffuse attenuation coefficient, upwelled radiance
DC Detritus absorption coefficient
DO Colored dissolved organic matter
DS Secchi depth
FA Fulvic acid concentration
HA Humic acid concentration
HS Salinity profiles
IM Inorganic suspended material concentration
OC Total dissolved organic carbon
OM Total dissolved organic matter
ON Particulate organic nitrogen
OS Organic suspended material concentration
PA Particle absorption coefficient
PB Phycobilipigment concentration
PC Phytoplankton species counts
PF Particle fluorescence
PH Phaeophytin concentration
PO Pigment concentration, other
PP Primary productivity
PR Photosynthetically available radiation (PAR)
PS Particle size spectra
RD Spectral sky radiance distribution
RF Rate of flow for along-track fluorescence
RS Spectral (remote sensing) reflectance
RT Rate of flow for along-track transmittance
SA Surface chlorophyll a concentration
SI Incident surface spectral irradiance
SP Surface phaeophytin concentration
ST Sea surface temperature
SU Surface upwelling radiance (extrapolated)
SW Surface wave height

TC Transmittance
TF Along-track fluorescence
TI Along-track incident (surface) spectral irradiance
TM Total suspended material
TP Temperature profiles
TR Spectral solar atmospheric transmission
TS Along-track salinity
TT Along-track sea surface temperature
VS Spectral volume scattering function
WD Wind direction
WR Water-leaving radiance (extrapolated)
WS Wind speed
WZ Water column depth
ZD Downwelled subsurface spectral irradiance
ZR Upwelled subsurface spectral radiance
ZU Upwelled subsurface spectral irradiance

2.2.4 Condensed Type Code

The condensed data type, T, has two values: 0 for cast data (i.e., vertical profiles) and 1 for all other data types (e.g., along-track, fixed mooring, drifters). For the former, NN is interpreted as a sequence number; and for the latter, NN is interpreted as an hour indicator, as described below.

2.2.5 Condensed Sequence Code

Interpretation of the condensed sequence code, NN, depends on the type code, T. If T=0, the sequence code is applied to cast data and is interpreted as a sequential cast number; if T=1, the sequence code is applied to non-cast data and is interpreted as a starting time of day indicator. For cast data, each column in NN is in the range 0,...,Z. The value 01 represents the first value in the sequence, 09

represents the ninth value, OZ represents the thirty-fifth, 10 represents the thirty-sixth, etc. A value of 00 is assigned for all sequences greater than or equal to 36×36, or 1,296.

For non-cast data, the left-most value in NN represents the starting hour indicator, in the range A,...,X, where A is 0 Greenwich Mean Time (GMT) and X is 23 GMT; the right-most value in NN represents the 2-minute indicator for data start, in the range 0,...,T, where 0 is for minute values of 0 and 1 past the hour, 1 is for 2 and 3 minutes past the hour, and T is for 58 and 59 minutes past the hour.

2.3 EXPANDED CONVENTION

The expanded file name uses the 0,...,Z character set of the condensed file names and adds one more: the hyphen (-) which is used as a sub-field embedded separator. The underscore (_) is used for undetermined or unknown elements, and the period (.) is used to separate the expanded name into smaller elements. Every expanded file name uses 24 characters starting with 1) a seven-character time stamp that identifies when the data was taken, 2) followed by a three-character position code identifying where the data was (geographically) taken, 3) then an eight-character code giving the measurement made, and finally 4) a three-character sequence identifying the type of data and its sequence. The expanded file naming convention is as follows:

YYMMMDD.PPP.IIIIIIII.TNN

where, YYMMMDD is a (Gregorian month) time stamp (YY is the decadal year with 50 representing 1950 and 49 representing 2049, MMM is an abbreviated month name which uses the first three characters of the Gregorian month, and DD is the day of the month); PPP is the position code (see Section 2.2.2); IIIIIIII is an expanded data identifier described below; T is the type code, and NN is the sequence code. (The latter two codes, T and NN, are interpreted as described earlier in Sections 2.2.4 and 2.2.5.) Purposely there is no definitive file extension associated with the expanded file name, so the user can append one based on local requirements.

The expanded identifier codes with the corresponding verbose descriptions are as follows:

MOCE01CR	MOCE-I cruise data
MOCE02CR	MOCE-II cruise data
Marra-WJ	Marra WOCE/JGOFS data
BBOP-Opt	Siegel BBOP data
BBOP-Pig	Michaels BATS pigment data
CHORSopt	CHORS optical data
JGOFS-EP	JGOFS EqPac data
NEToptic	NET optical data
NETpigmt	NET pigment data
AirFluor	Airborne fluorescence
AirHumid	Air relative humidity
AirDnIrr	Airborne spectral downwelled irradiance
Atm-Pres	Atmospheric pressure
AirUpRad	Airborne spectral upwelled radiance
Air-Temp	Air temperature
AT-Fluor	Along-track fluorescence
AT-Salin	Along-track salinity
ATIncIrr	Along-track incident (surface) spectral irradiance
ATSfcTem	Along-track sea surface temperature
BeamAten	Beam attenuation coefficient
ChlaFlPr	Chlorophyll a fluorescence profiles
ChlaProf	Chlorophyll a profiles
CoccoCon	Coccolith concentration
ColorDOM	Colored dissolved organic matter
DetAbsrp	Detritus absorption coefficient
DifDnIrr	Diffuse attenuation coefficient, downwelled irradiance
DifUpRad	Diffuse attenuation coefficient, upwelled irradiance
FulvAcid	Fulvic acid concentration
HumiAcid	Humic acid concentration
InorgSMC	Inorganic suspended material concentration
LWRadExt	Water-leaving radiance (extrapolated)
OSM-Conc	Organic suspended material concentration
Part-Abs	Particle absorption coefficient
ParticFl	Particle fluorescence
ParticON	Particulate organic nitrogen
PartSpec	Particle size spectra
PhaeoCon	Phaeophytin concentration
PhotoRad	Photosynthetically available radiation (PAR)
PhycoCon	Phycobilipigment concentration
PhytoSpe	Phytoplankton species counts
PigConOt	Pigment concentration, other
PrimProd	Primary productivity
ROF-ATFl	Rate of flow for along-track fluorescence
ROF-ATTr	Rate of flow for along-track transmittance
Sal-Prof	Salinity profiles
Sat-Chla	Satellite pigment (chlorophyll a) concentration (computed)
SecDepth	Secchi depth
Sfc-Chla	Surface chlorophyll a concentration
Sfc-Temp	Sea surface temperature
SfcPhaeo	Surface phaeophytin concentration
SfcSpIrr	Incident surface spectral irradiance
SfcUpRad	Surface upwelling radiance (extrapolated)
SfcWavHt	Surface wave height
SolTrans	Spectral solar atmospheric transmission
Spec-Abs	Spectral absorption coefficient

SpecAten	Spectral beam attenuation coefficient
SpecBack	Spectral backscattering coefficient
SpecRefl	Spectral (remote sensing) reflectance
SpSkyRad	Spectral sky radiance distribution
SubDnIrr	Downwelled subsurface spectral irradiance
SubUpIrr	Upwelled subsurface spectral irradiance
SubUpRad	Upwelled subsurface spectral radiance
TempProf	Temperature profiles
TotalDOC	Total dissolved organic carbon
TotalDOM	Total dissolved organic matter
Total-SM	Total suspended material
Transmit	Transmittance
Vol-Scat	Spectral volume scattering function
WatDepth	Water column depth
Wind-Dir	Wind direction
Wind-Spd	Wind speed

2.4 ARCHIVING PROCEDURES

After preparation of the file headers is complete, the files are archived to the SeaWiFS calibration and validation QStar magneto-optical (MO) jukebox. Science Team Members (STMs) and other authorized users will be able to query the contents of the archive through a GUI built with IDL software and interfaced to the Sybase tables (Fig. 3).

To access the GUI, the user must provide a valid Sybase server account name and password. This provides an extra measure of security to supplement the security at the operating system level governing the user's initial entry to node calval.gsfc.nasa.gov. Once within the GUI, the user specifies a particular data set class to query (i.e., SIRREX, SBRC, or bio-optical), and particular values for one or more of the required keywords for the chosen class (where appropriate, a pop-up window provides a list of valid values for the selected keywords, taken from the database).

After the user runs the query, the software polls one or more database tables in order to construct a list of files which meet the stated criteria (these are printed in a separate window as shown in Fig. 3). The user can then select one or more of the listed files, in order to either request the data (and optionally, the catalog of headers and comments and external documents from the data provider, if applicable) or to view the first few records of the selected files in a separate GUI window. The windows containing the file names and file excerpts are scrollable in both the horizontal and vertical directions.

In addition to making queries, the users can also revise their address, phone number, and other personal data in the database, or leave a comment for the system administrator. Each request results in a record being entered in a special Sybase table. In addition, an electronic mail (e-mail) message is sent to the bio-optical system manager and a verification request (in the form of a letter or telephone call, depending on the user's preference) is sent to

the user. When the response to the verification request is received, it is also logged into the database and the data can then be retrieved by the user. Choices for file retrieval include network or various types of tape, including 9-track, 8 mm (EXABYTE), and 4 mm digital audio tape (DAT). When network transfers are requested, files will be staged from the MO to the user's account for file transfer protocol (ftp) access. With the data, an ASCII look-up table will be provided which lists gross characteristics of the data (i.e., the same ones covered in the required keyword set).

Each user needing access to the archive sends an e-mail message to seabass@calval.gsfc.nasa.gov in order to receive a database user name and password and further instructions. All of the users outside of GSFC will be placed in a common group (both at the UNIX level and the Sybase level) to control system and database access. A series of administrative tables present in Sybase will track the users, their access to the system, their requests for data, and their comments on the system (see Table 1). Requests for data will be monitored by the administrator of the bio-optical data system, who will also be responsible for copying the requested data to disk or tape, once verification is received from the user.

The proposed breakdown of tasks and milestones within the bio-optical data system is detailed in Table 3. A group effort will be required to accomplish all of the necessary tasks before launch of the SeaWiFS instrument. E-n. Yeh is tasked with handling the initial preparation of most of the data that is not derived from SeaWiFS activities including renaming, restructuring if necessary, and the insertion of header information. T. Westphal and Y. Ge performed similar functions for the SeaWiFS data sets, including SIRREX activities, MOCE cruises, and MOBY deployments. J. Firestone handles applications development, including creation, population, and maintenance of the Sybase tables, jukebox applications, and development of the user GUI interfacing with the database.

2.4.1 Archived File Example

An example data file, BQUINZOW.001, with a complete header block (presented with a truncated /experiment entry to save on space), and an abbreviated data block, is presented below:

```
/begin_header
/affiliations=Lamont-Doherty_Earth_Observatory
/column_headers=Marra
/data_file_name=bquinz0w.001
/data_type=CAST
/documentation_file_name=README,statlog,
oceanmsg1,PCparbat,dtbat
/end_date=1991SEP04
/end_latitude=-12.013[DEG]
/end_longitude=-151.021[DEG]
/end_time=22:21:00.000[GMT]
/experiment=JGOFS-WOCE_Hydrographic_Program
```

Table 3. Bio-optical data system tasks and milestones.

Brief Description of Activities	1993			1994								
	O	N	D	J	F	M	A	M	J	J	A	S
Database design.	■											
Implement GUI prototype.	■											
SIRREX header creation and metadata archiving.	■											
Header creation and metadata archiving for the CTMS† data sets.			■									
Develop C routines for file renaming.			■									
First Bio-optical Algorithm working group meeting.			■									
MOCE-I master log creation.												
Data system overview and historical pigment database documentation.						■						
Implement operational GUI.							■					
MOCE header creation and metadata archiving.							■					
Second Bio-optical Algorithm Working Group Meeting.								■				
Archive all data to the jukebox.									■			
User test period.										■		
System enhancements.										■		
SIRREX-3 at CHORS, San Diego.												■

† Data sets developed by D. Clark, C. Trees, J. Marra, and D. Siegel (CTMS)
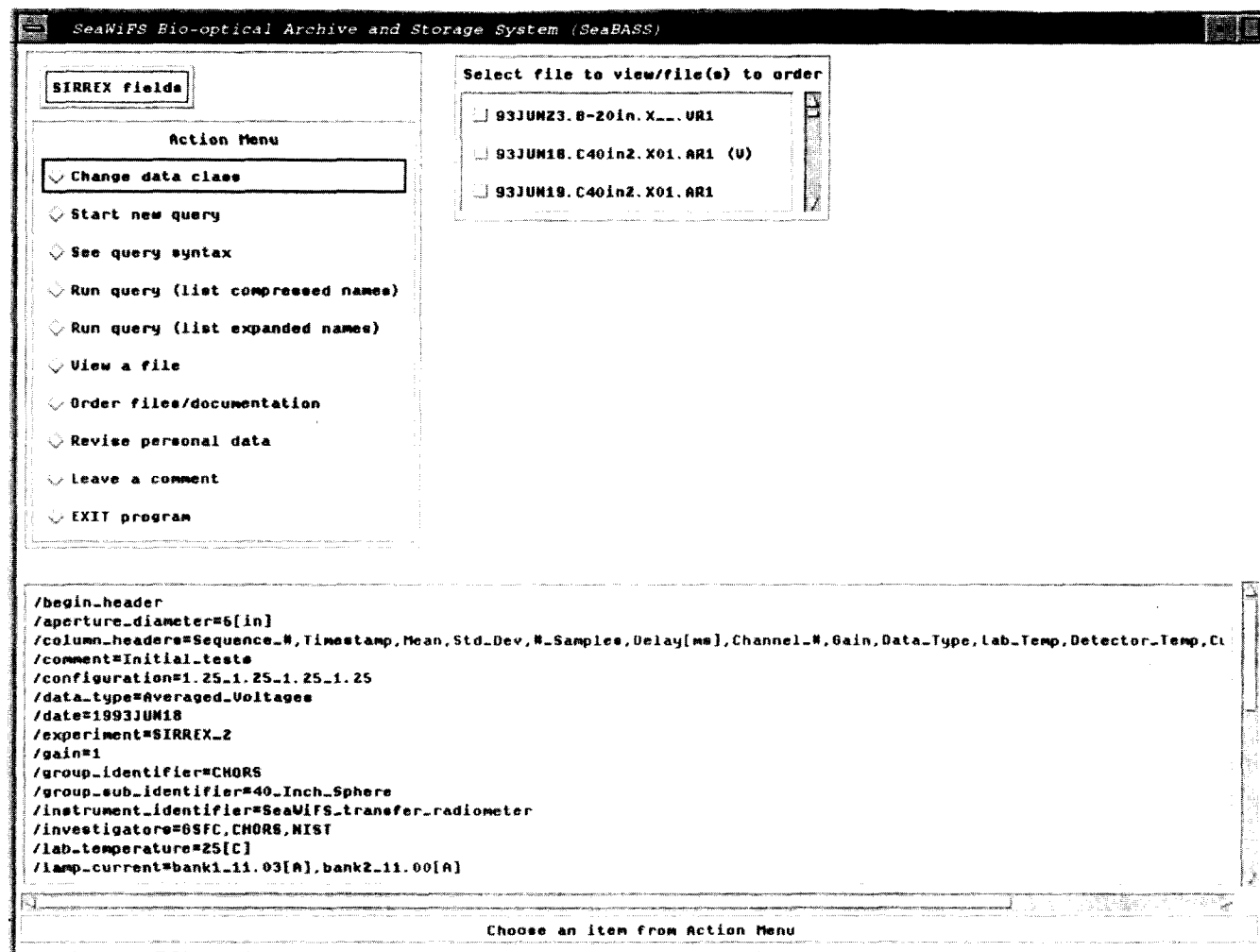
Fig. 3. Prototype GUI used to access information on bio-optical data files.

```
/investigators=John_Marra,Carol_Knudson
/original_file_name=p910904a.dec
/sequence_number=1
/start_date=1991SEP04
/start_latitude=-12.013[DEG]
/start_longitude=-151.021[DEG]
/start_time=22:21:00.000[GMT]
/end_header@
3.000000 1.295326e+02
4.000000 1.302285e+02
5.000000 1.291834e+02
    :
193.000000 1.355593e+02
194.000000 1.357831e+02
195.000000 1.357619e+02
```

2.4.2 Condensed File Name Example

The sample file name BQUINZOW.001 represents the following:

Date BQU = 4 September 1991,
Position INZ = grid box having its northwest corner at (−152.5,−10),
Data identifier OW = Marra WOCE/JGOFS data,
Type code 0 = cast data, and
Sequence and time code 01 = first cast in the sequence.

It is worth noting again that a verbose description of the file name is included in the keywords:

```
/column_headers,
/data_type,
/sequence_number,
/start_date,
/start_latitude,
/start_longitude, and
/start_time.
```

2.4.3 Expanded File Name Example

Using the same sample file, the date stamp of 4 September 1991 would be coded as 91Sep04. The position code remains unchanged, INZ. The data identifier is written out with its expanded representation, Marra-WJ. The sequence and time code remains unchanged, 001. Hence the full expanded file name is coded as:

91Sep04.INZ.Marra-WJ.001.

The expanded format uses only 24 of the 31 characters normally available to many operating systems capable of extended file naming. The remaining characters are available for use in local applications, perhaps for tracking post-processing of the data.

2.5 KEYWORD GLOSSARY

The keyword glossary is composed of two sets for bio-optical (excluding SIRREX) data files: required and approved keywords. The former must appear in every data file, while the latter appear only as needed to fully describe the contents of the file. Of course, local use of a data file permits the addition of locally defined keywords, perhaps for tracking post-processing of the data.

2.5.1 Required Keywords

The required keywords for the bio-optical data (which excludes SIRREX data) are as follows:

/begin_header This keyword does not have a value and must be the first entry in the series of lines comprising the header block.

/affiliations⟨s⟩ This is a comma-delimited list of affiliations for individuals included in /investigators. All spaces in a given affiliation should be filled with underscores.

/column_headers=⟨s⟩ This is a comma-delimited list of headers for the columnar data (parameters) in the input file. The ith entry in the list is associated with the ith column. Units for the parameters should be included, enclosed in brackets. All spaces in a given column header should be filled with underscores. For standardization, the individual entries should be the *short names* defined in Section 2.2.3.

/data_file_name=⟨s⟩ This is the name of the output file using the condensed naming convention described in Section 2.2.

/data_type=⟨s⟩ The allowed data types are: ALONG_TRACK, CAST, DRIFTER, or MOORING.

/end_date=$yyyymmmdd$ The ending date of the observations. In this format, $yyyy$ is the year, mmm is the abbreviated three-character Gregorian month of the year, and dd is the day of the Gregorian month. The month abbreviations use the first three characters of the month: Jan is January, Feb is February, etc.

/end_latitude=⟨n⟩[u] This is the ending latitudinal position where the data was collected. The allowed values are -90 to 90, with degrees north (of the equator) positive.

/end_longitude=⟨n⟩[u] This is the ending longitudinal position where the data was collected. The allowed values are -180 to 180, with degrees east (of the prime meridian) positive.

/end_time=⟨s⟩[u] This is the ending time of the observations. The format for ⟨s⟩ is $hh:mm:ss.sss$, where hh is the 24-hour hour of the day, mm is the minute of the hour, and $ss.sss$ is the decimal second of the hour. The $ss.sss$ field is optional and the [u] field is used to denote GMT, LOC, EST, EDT, etc., nominally [GMT]. For example, 12:05:23.300[LOC] indicates 12:05 and 23.300 seconds local time, while 12:05:23.300[GMT] indicates the time is GMT (i.e., relative to 0° longitude).

/experiment=⟨s⟩ This is the formal name of the experiment for which data was collected. All spaces in a given experiment name must be filled with underscores.

/investigators=⟨s⟩ This is a comma-delimited list of the persons who collected the data. All spaces in a given investigator name must be filled with underscores.

/original_file_name=⟨s⟩ This is the name of the original file submitted to the Project, as originally named by the investigator(s).

/start_date=yyyymmmdd The starting date of the observations. In this format, yyyy is the year, mmm is the abbreviated three-character Gregorian month of the year, and dd is the day of the Gregorian month. The month abbreviations use the first three characters of the month: Jan is January, Feb is February, etc.

/start_latitude=⟨n⟩[u] This is the starting latitudinal position where the data was collected. The allowed values are -90 to 90, with degrees north (of the equator) positive.

/start_longitude=⟨n⟩[u] This is the starting longitudinal position where the data was collected. The allowed values are -180 to 180, with degrees east (of the prime meridian) positive.

/start_time=⟨s⟩[u] The starting time of the observations. The format for ⟨s⟩ is hh:mm:ss.sss, where hh is the 24-hour hour of the day, mm is the minute of the hour, and ss.sss is the decimal second of the hour. The ss.sss field is optional and the [u] field is used to denote GMT, LOC, EST, EDT, etc., nominally [GMT]. For example, 12:05:23.300[LOC] indicates 12:05 and 23.300 seconds local time, while 12:05:23.300[GMT] indicates the time is Greenwich Mean Time (relative to 0° longitude).

/sequence_number=⟨n⟩ This is a counter, used only for cast data; this is not needed for other data types.

/end_header@ This keyword does not have a value and must be the last entry in the series of lines comprising the header block. Note the at-sign (@) at the end.

There is no specific requirement regarding the ordering of the keywords in the files, other than the /begin_header and /end_header@ keywords, which must be first and last respectively. It is recommended, however, that the second through next to last keywords be kept alphabetical to maximize the readability of the file.

2.5.2 Approved Keywords

The additional approved keywords for bio-optical data files are as follows:

/documentation_file_name=⟨s⟩ This is a list, which is comma-delimited, of associated documentation files. All spaces in a given documentation file name should be filled with underscores.

2.6 DISCUSSION

The bio-optical data archive contains files stored in an ASCII text format that is independent of platform. Each file has a header block containing information regarding when, where, how, and by whom the data was taken. This header also contains information specifying how many rows and columns of data to expect. Headers from original data files are placed into bio-optical data keywords, when appropriate. A sequence column is added if necessary. Each column of data is written in a single numeric format before the file is placed in the archive. The result is a file that is useable on many platforms and has descriptive header information in a known format that is easily parsed for automated processing. This header contains the details of how the data was obtained, as well as how much and what type of data is contained in the file.

Chapter 3

The Historical Pigment Database

JAMES K. FIRESTONE
General Sciences Corporation, Laurel, Maryland

CHARLES R. MCCLAIN
NASA Goddard Space Flight Center, Greenbelt, Maryland

ABSTRACT

The SeaWiFS calibration and validation historical pigment database has been assembled for evaluating bio-optical algorithms, and the quality control of derived products from SeaWiFS and other satellite ocean color sensors. As of this writing, there are more than 330,000 independent pigment observations in the database, representing over 50 data sets amassed between 1956 and 1994 from a wide variety of investigators and oceanographic expeditions. This document describes the structure of the database, the current set of holdings, the procedures for submitting pigment data to the SeaWiFS Project, and the methods for converting data contributions from the data format in which they are submitted to the database format.

3.1 INTRODUCTION

Figure 4 depicts the historical pigment data in the context of the overall SeaWiFS bio-optical data system. The data arrives in numerous formats, although it is usually stored as either ASCII files or as spreadsheets from commercial software packages. As a first step, plots of the geolocation of each station are used to check the validity of the latitude and longitude entries. In a few cases, a station may be displayed over land because of an error in the coding of latitude or longitude. In these cases, the associated records are not imported into the database—they are rejected. A high-level check is also made on the magnitude of the pigment values; if they appear out of range (usually impossibly large), the data is not included.

As additional QC checks are defined by the Bio-Optical Algorithms Subgroup (Bio-Optical Algorithms in Fig. 4) of the SeaWiFS Science Team (Hooker et al. 1993), they will be incorporated into the processing of new data sets and will be run retroactively on the existing archive. After the QC checks are completed, the data is made compatible with the database structure, as discussed in Section 3.3. Once the data is loaded into the database, various statistics and secondary products are derived as required by the Bio-Optical Algorithms Subgroup. Again, the Bio-Optical Algorithms Subgroup will be advising the CVT on the specific products and statistics to store in the database (SeaBASS in Fig. 4). The results are made available, on request, to members of the Science Team and authorized database users, and will be used in the ocean color product verification activity within the SeaWiFS Calibration and Validation Program (Fig. 4).

At some future date, the contents of the database may also be made publicly available through the DAAC. Currently, the data can be easily listed and displayed through the JGOFS object-oriented data system, as described in Section 3.5. In Fig. 4, everything within the dashed box is the responsibility of CVT personnel. The CVT also participates in the Bio-Optical Algorithm Subgroup of the SeaWiFS Science Team.

3.2 PIGMENT DATABASES

The CVT is maintaining four separate data sets containing the entire historical pigment database. Two of the data sets are resident in the Microsoft FoxPro environment on a 486 PC system, and are stored as standard .DBF files. These files are, therefore, accessible from other PC database packages, such as Borland's dBase IV. One of the PC data sets contains only near-surface pigment values at depths between 0–10 m, while a second contains pigment profiles, i.e., no depths excluded. There is an equivalent set of two data sets, with identical structures and data to the PC data sets, but stored under the Sybase relational database management system (RDBMS) running on a Silicon Graphics, Inc. (SGI) UNIX workstation. Table 4 shows the database schema for the four historical pigment data sets and sample records.
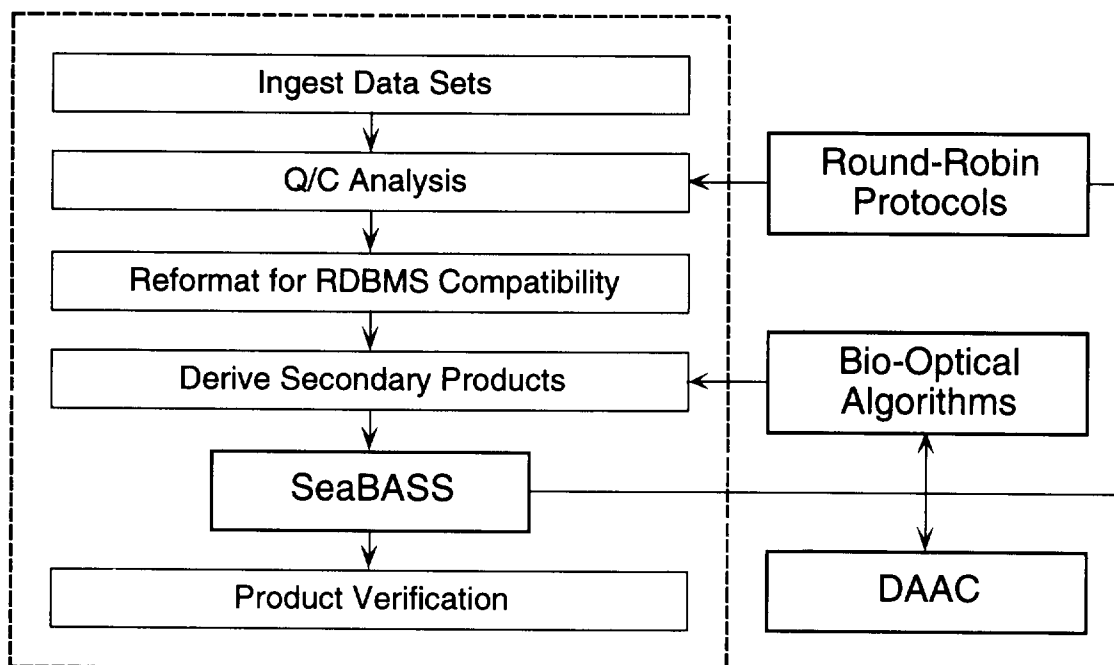
Fig. 4. The historical pigment data in the context of the overall data system.

Table 4. Database schema for the historical pig-
ment data sets. In the *Format* column: i, a, and f
indicate an integer, character, and a real field, re-
spectively; the numbers indicate the width of the
fields using a standard programming notation.

No.	Variable	Description	Format
1	Year	4-digit	$i4$
2	Month	1–12	$i2$
3	Day	1–31	$i2$
4	Hour	0–23	$i2$
5	Minute	0–59	$i2$
6	Second	0–59	$i2$
7	Time_Zone	GMT, etc. 1	$a3$
8	Latitude	−90° to 90° (S–N)	$f8.4$
9	Longitude	−180° to 180° (W–E)	$f9.4$
10	Depth	meters (≥ 0)	$f8.2$
11	Source	Source Tag 2	$a20$
12	Cruise	Cruise Name 3	$a20$
13	Station	Station Name 3	$a20$
14	Chlor_a	Chl. a 3 4 5	$f6.3$
15	HPLC_Chla	HPLC Chl. a 3 4	$f6.3$
16	Chlor_b	Chl. b 3 4	$f6.3$
17	Chlor_c	Chl. c 3 4	$f6.3$
18	Total_pig	Total Pigment 3 4	$f6.3$
19	Phaeopig	Phaeopigment 3 4	$f6.3$
20	Chlapphaeo	Chl. a + Phaeo. 3 4	$f6.3$
21	Comment	Useful Information	$a80$

1 Blank if missing.
2 Submitter's name, data location, or original file name.
3 Specify -9999 if missing.
4 Units of $mg\,m^{-3}$.
5 Derived from non-HPLC sources, e.g., Turner method.

3.3 DATA PROCESSING

The goal of this activity is to convert the data received
from outside investigators to a standard form directly im-
portable into the four databases described above. Since
the data comes to the Project in various physical forms
(generally ASCII or spreadsheet) and with different inter-
nal structures, considerable effort is sometimes needed to
restructure it.

The most common problems are encountered when at-
tempts are made to read vastly different file formats and
data structures. These problems are as follows:

1. *Physical structure differences:* typically, the da-
   ta may be in either an ASCII file or a spread-
   sheet file. Spreadsheets may be output from dif-
   ferent software packages, such as Lotus 1-2-3 or
   Microsoft Excel.

2. *Position representation discrepancies:* dissimi-
   larities here include sign conventions, primarily
   in longitude, indicating the hemisphere where
   the data is located; use of directional indicators,
   such as N for north and W for west to replace
   signs; and use of degree/minute/second repre-
   sentations versus the use of decimal degrees.

3. *Non-standard dates and times:* for date, some
   investigators have chosen to use the sequential
   day of the year while others have chosen the Gre-
   gorian month and day. The month has been ex-
   pressed as abbreviations (e.g., Jan for January)
   or numerically (e.g., 1 for January and 12 for De-
   cember). The time may be expressed as GMT

or as local time, and often there is no indication of which one is being used. Hour and minute values may be colon-separated or stored in separate columns in the file.

4. *Field ordering:* each investigator uses his or her own method for writing the data files. Therefore, it is generally necessary to reorganize, and in some cases add or remove, columns from the data so it conforms with the established output standard.

5. *Missing data:* if a data value is missing for a particular column, it may be left blank or filled with almost any invalid data value as a replacement. When left blank, data reshuffling is made more difficult because most spreadsheets will move leftward all columns to the right of the missing columns when data is imported.

6. *Field delimiter differences:* various types of field delimiters have been used, most commonly tabs, spaces, and commas.

Typically, one of three approaches has been used in generating a tab-delimited version of the fields listed in Table 4 (tab-delimited ASCII files can be easily imported into COTS programs). The three approaches are as follows: a) use the Freeform software package provided free of charge by the National Oceanic and Atmospheric Administration (NOAA) National Geophysical Data Center (Habermann 1991), b) use a commercial spreadsheet such as Lotus 1-2-3 or Microsoft Excel, or c) write a customized program, to reorder the columns or compute standardized representations for spatial and temporal parameters.

The Freeform package, the first approach, allows specification of both the input and desired output file formats in an ASCII text file, and eliminates the need for writing a custom program to unpack each unique input format. It also supports conversions between different representations for space and time. However, this approach works only if the input data is in a fixed set of columns and never varies, since the start and end positions of each column must be specified. In addition, headers can be read only if they are found at the beginning of the file. Nevertheless, Freeform has been useful for processing many of the more straightforward data sets.

Commercial spreadsheets, the second approach, are extremely useful when the input data are not in uniform columns, but the columns are delimited in some fashion. However, as described above, they have difficulty aligning data when blanks have been used as the missing value.

Customized programs, the third approach, are a last resort and the most time consuming, since testing of the code is involved for each input format. This approach is sometimes the only choice when the input cannot be processed with either of the other approaches. For example, headers may be interleaved with the data. This sometimes happens with station data where the header gives the location, time, and perhaps cruise information.

The output of any of the discussed approaches is a tab-delimited ASCII file in the format of Table 4. This file is then easily imported into FoxPro (using the *Append From...* menu option) or Sybase (using the bulk-copy, or bcp, utility).

3.4 DATA HOLDINGS

Table 5 is a list of the current set of data holdings in the historical pigment archive. Column 2 lists the institutional, or individual, source of the data (entries ingested since the fall of 1993 are followed by a † symbol). Column 3 indicates the name of the provider's original file, in either ASCII or spreadsheet format. Column 4 indicates the general location of the data set. Columns 5 and 6 represent the number of unique spatial points (i.e., one per unique location, even over multiple times and depths at that location), broken down by data source. The values in column 5 are computed independently for each data source, without regard to whether the data are present elsewhere in the other data sets. The values in column 6 are computed over the entire archive, then broken down by source, so that no duplicates are found between data sets. In cases where a data source contains data found elsewhere in the archive, e.g., W. Balch (University of Miami/Rosenstiel School for Marine and Atmospheric Sciences data), column 6 values will be smaller than those in column 5. Both columns 5 and 6 will be much smaller than the total number of observations in a data set when the data set is largely composed of a time series (e.g., the JGOFS North Atlantic bloom data is a time series, so only one unique location is listed in columns 5 and 6).

Figure 5 is a depiction of all data locations in the historical pigment archive plotted on a world map. A program called PLOTPNT within the UNIX version of the SEA-PAK (McClain et al. 1991a and 1991b) software system was used to plot the locations, using the latitude and longitude columns in each data file as input. The area least sampled continues to be the southern ocean, particularly the Atlantic and Indian Ocean portions.

3.5 ACCESSING THE ARCHIVE

In order to make the pigment archive readily available to the scientific community, the CVT has provided the data to the JGOFS Distributed Object Oriented Data System. This system, developed and overseen by G. Flierl of the Massachusetts Institute of Technology (MIT), D. Glover of Woods Hole Oceanographic Institution (WHOI), and J. Bishop and S. Paranjpe of Lamont-Doherty Earth Observatory (LDEO), makes widely scattered data sets available to anyone having the JGOFS server installed on their Sun Microsystems (Sun), Digital Equipment Corporation (DEC), SGI, International Business Machines (IBM) or similar UNIX workstations. With the software, data objects can be easily listed or displayed in a variety of ways,

Fig. 5. Map depicting locations of historical pigment data holdings as of June 1994.

Table 5. Historical pigment data sets in the archive. Column 5 indicates the number of unique spatial points for each data source. Column 6 indicates the number of unique spatial points over the entire data set for each source.

Entry No.	Data Source	Original File Name	Geolocation Description	Unique Stations in Source	Database
1	French	7906F.WK1	Equatorial Atlantic	29	29
2	French	7910F.WK1	Equatorial Atlantic	49	49
3	French	7912F.WK1	Equatorial Atlantic	39	39
4	French	7992F.WK1	Equatorial Atlantic	9	9
5	French	7996F.WK1	Equatorial Atlantic	9	9
6	French	8001F.WK1	Equatorial Atlantic	17	17
7	French	8201F.WK1	Equatorial Atlantic	46	46
8	French	8301F.WK1	Equatorial Atlantic	68	68
9	French	8303F.WK1	Equatorial Atlantic	92	92
10	French	8401F.WK1	Equatorial Atlantic	94	94
11	French	8403F.WK1	Equatorial Atlantic	99	99
12	French	8601F.WK1	Equatorial Atlantic	136	136
13	Dandonneau	Tri81	Equatorial Pacific	1,898	1,898
14	Dandonneau	Tri82	Equatorial Pacific	4,464	4,464
15	Dandonneau	Tri83	Equatorial Pacific	4,739	4,739
16	Dandonneau	Tri84	Equatorial Pacific	3,505	3,505
17	Dandonneau	Tri85	Equatorial Pacific	3,773	3,773
18	Dandonneau	Tri86	Equatorial Pacific	3,333	3,333
19	Chavez	CHAVEZ.TXT	Equatorial Pacific	500	500
20	JARE§	J74_75.TXT	Pacific	176	176
21	JARE	J75_76.TXT	Pacific	113	113
22	JARE	J76_77.TXT	Pacific	150	150
23	JARE	J77_78.TXT	Pacific	143	143
24	JARE	J78_79.TXT	Pacific	241	241
25	JARE	J79_80.TXT	Pacific	216	216
26	JARE	J80_81.TXT	Pacific	162	162
27	JARE	J81_82.TXT	Pacific	135	135
28	JARE	DATA7-27.PRN	Pacific	1,496	1,496
29	Oregon State U.	OSU*.ASC	Pacific	619	619
30	Russian/Sushkina	RUSSBIOL.ASC	Pacific	126	126
31	Balch†	*.DAT	Northern Hemisphere	8,669	7,946
32	Yentsch	YENTSCH.TXT	North Atlantic	2,460	2,460
33	Konovalov-Beljaeva	RUSSIAN.WK1	Atlantic	120	120
34	JGOFS	JGOFS.ASC	North Atlantic Bloom	1	1
35	Atkinson	BLMPIG.WK1	South Atlantic Bight	17	17
36	Yoder	PREGABPG.WK1	South Atlantic Bight	15	15
37	Yoder	GABEAPIG.WK1	South Atlantic Bight	21	21
38	Yoder	GABISLPG.WK1	South Atlantic Bight	12	12
39	Yoder	GULFPIG.WK1	Gulf of Mexico	20	20
40	Northeast Fisheries	MARMAP.TXT	Mid-Atlantic Bight	750	750
41	Muller-Karger	VENIZPIG.WK1	Caribbean Sea	21	21
42	Barale	ADRIAPIG.WK1	Adriatic Sea	31	31
43	Russian	RUSS22AV.ASC	Indian Ocean	158	158
44	Russian	RUSS08PV.ASC	Indian Ocean	109	109
45	Wright-Australia	AAMBER2P.RYD	Southern Ocean	46	46
46	Wright-Australia	ADBEXNOV.198	Southern Ocean	34	34
47	Wright-Australia	AAMBERTR.ANS	Southern Ocean	36	36
48	Wright-Australia	HEARDISJ.UNE	Southern Ocean	61	61
49	Wright-Australia	SIBEX2PR.YDZ	Southern Ocean	46	45

§Japanese Antarctic Research Expedition   * File name wildcard specifier.   † Recent update to the database.

Table 5. (cont.) Historical pigment data sets in the archive. Column 5 indicates the number of unique spatial points for each data source. Column 6 indicates the number of unique spatial points over the entire data set for each source.

Entry No.	Data Source	Original File Name	Geolocation Description	Unique Stations in Source	Database
50	Wright-Australia	WOCE1-OCT.19	Southern Ocean	27	27
51	Dieckmann/Germany	WEDDELL*.TXT	Weddell Sea	82	74
52	Dieckmann/Germany	ANTV_*.WK3	Weddell Sea	57	57
53	Sullivan	AM123CHL.WKS	Weddell Sea	244	244
54	El-Sayed	ELSAY*.WK1	Southern Ocean	397	397
55	British	BIOMASS*.TXT	Southern Ocean	474	474
56	New Zealand	KIWI1978.TXT	Southern Ocean	52	52
57	Sakshaug/Holm-Hansen	POLSEAII.ASC	Southern Ocean	207	207
58	Longhurst†	BEDFORD.TXT	North Atlantic	1,885	1,885
59	Biggs†	BIGGSPRO.FIL	Southern Ocean	34	34
60	Clark Along-Track†	*.TLK	United States Coastal	8,566	8,566
61	Clark Stations†	*_NOAA_*.DAT	United States Coastal	62	62
62	Mitchell†	R2CHL.ASC	Southern Ocean	345	69
63	JGOFS EqPac†	PIGT*.PRN	Equatorial Pacific	154	154
64	Marra JGOFS†	P16CHL.ASC	Equatorial Pacific	25	25
65	Siegel BBOP†	E-mail messages	Bermuda	28	28
66	Michaels BATS†	PIG*.DAT	Bermuda	17	17
			Total Number of Stations	51,759	50,751

† Recent update to the database.　　　　* File name wildcard specifier.　　　　Current as of June 1994.

and certain derived quantities can also be included in the output. J. Bishop of LDEO has written a *method* (data reading program) to accompany the data. Together the method and the data form a data *object* which is accessible from the JGOFS software. The GSFC pigment data is found in the following object:

```
name::pigsn(/geochem/solar/data/pigs/pigsfc.asc)
```

where the internet protocol (IP) address of the system name is 129.236.40.87. The CVT intends to provide periodic updates of the archive to the JGOFS system. For a complete description of the design, installation procedures and use of the JGOFS software, consult Flierl et al. (1993). A capsule summary of the installation procedure taken from this document is found in Section 3.7. The contents of Flierl et al. (1993) can be found in the doc subdirectory (after installation) as a series of .ps (PostScript) or .tex (TeX) files for hardcopy output.

3.6 SUBMISSION GUIDELINES

The CVT encourages submission of historical pigment data sets collected from specialized field expeditions or as part of coordinated global programs such as JGOFS or WOCE. The general guidelines for data submission are detailed in Chapter 1. In addition, there are two additional guidelines which apply specifically to the pigment data.

1. *Record format:* It is particularly important to fill each column for the data records, even if data

is missing. The use of blanks for missing data can make it more difficult to import the data into the database. In lieu of blanks for missing data, a code of -9999 must be used. Also, the content of data headers (e.g., station names, locations, times, etc.), when possible, should be converted to the columnar representation and included with the data rather than remaining text entries.

2. *Extra characters:* Do not include extra characters in the data, such as boolean operators. There is currently no provision to store this level of detail in the database.

3.7 JGOFS SOFTWARE

The procedures to install the JGOFS software are as follows:

1. Set up a jgofs account on a local machine to store the software.
2. Log in to the local jgofs account and telnet to IP address pimms.mit.edu (18.83.0.104).
3. Log in as the account jgofs, using the password object$data. The system asks for your machine name, possibly the IP address, and username (the name under which you would like to receive mail).
4. Select option 1 (transfer JGOFS software). The system executes an ftp command back to your

machine. Press a return if your local user name is `jgofs`; otherwise, enter the user name. Then enter the password.

5. Copy the software to your local machine by typing `put jgofs.tar` and then `quit`.

6. Exit from the control menu by typing 0.

7. On your local system, unpack the tar file by typing `tar -xvf jgofs.tar`.

8. Edit the file called `Makefile`. If you are not on a Sun workstation, comment out the lines for the Sun and activate the lines for your own system (a `#` sign at the start of a line indicates a comment).

9. Type `setenv PWD 'pwd'`.

10. Type `make`.

11. If this works, type `bin/jgofs_serv &` to start the server.

12. Then, initialize the system for plotting by typing `source penplot`.

13. Start the menuing system by typing the command `source ~jgofs/jgofs`. A three-window display, with windows labeled `xmenu`, `list`, and

jgofs should pop up.

14. Click in the bottom portion of the window entitled `xmenu`.

15. Test the software by typing `list "#test"`. A printout of hydrographic data retrieved from a local data object, which comes with the installation, should appear. Type q to quit.

16. Check the listing of remote objects by typing `list "#tco2"`.

17. If the menu system does not build properly, the local machine may be missing portions of the full X library. In this case, Glenn Flierl should be contacted at either `glenn@lake.mit.edu` or `glenn@pimms.mit.edu` for advice.

18. Normally, the JGOFS server, once started, will run until the next system reboot. If the system reboots, the user will need to manually start the JGOFS server with step 11 above (or you can add it to your system startup procedure in `/etc/rc.local`. To stop the server at any time, type `source bin/stopserver` from your local jgofs account.

Chapter 4

The SIRREX Database

STANFORD B. HOOKER

NASA Goddard Space Flight Center, Greenbelt, Maryland

TODD L. WESTPHAL

General Sciences Corporation, Laurel, Maryland

YUNTAO GE

University Space Research Association, Columbia, Maryland

ABSTRACT

This document describes the procedures used in obtaining and archiving the SIRREX data sets. These data sets provide a calibration history for the various lamps, spheres, plaques, and radiometers used as a radiometric standard for field measurements. Currently, there are two SIRREX data sets in the archive. This document contains the file naming and header protocol used for SIRREX data, the procedure for archiving the original data, and an example of an archived file.

4.1 INTRODUCTION

In keeping with the SeaBASS file architecture, SIRREX data utilizes a condensed and expanded file naming convention. The former makes use of the so-called *eight-point-three* 11-character format designed for PCs running MS-DOS (the most restrictive naming convention commonly in use); the latter is for any computer supporting 31-character file names. All files in the database are stored in the condensed format, but users can request data in the expanded format which is more readily deciphered.

4.2 CONDENSED CONVENTION

Every condensed file name starts with a three-character time stamp that identifies when the data was taken, followed by a five-character primary test code identifying what was measured and with what instrumentation, and then a three-character extension identifying the type of data recorded. The condensed file naming convention is as follows:

DDDPPPPP.eee

where DDD is the date stamp, PPPPP is the primary test code, and eee is the file extension. The date stamp is composed of three characters: YMD where, Y is the year code, M is the month code, and D is the day of the month code. The primary test code is composed of three sub-codes: SSICC where, SS is a two-character source identifier, I is a one-character measurement instrument identifier, and CC is a

two-character configuration identifier. The file extension is also further divided into three codes: drs where, d is a one-character data type, r is a one-character run type, and s is a one-character sequential run character (frequently a numeric sequence). Thus, the complete condensed file name convention is:

YMDSSICC.drs

where, once again, the condensed convention makes full use of the *eight-point-three* 11-character format designed for MS-DOS PCs.

4.2.1 Condensed Date Code

The condensed time code is a three-character code, YMD, and is decoded as follows: Y has the allowed values $1, \ldots, Z$, where 1=1991, 2=1992, 3=1993, A=2000, and Z=2025; M has the allowed values $1, 2, 3, \ldots, A, B, C$, where 1=Jan., 2=Feb., 3=Mar., A=Oct., B=Nov., and C=Dec.; and D has the allowed values $1, 2, 3, \ldots, A, B, \ldots, V$, where 1=1, 2=2, 3=3, A=10, B=11, and V=31. If any part of the date field is not known, the underscore (_) is used for the unknown component.

4.2.2 Source Identifier Code

The source identifier code is a two-character code which identifies what radiance source was employed, usually a sphere, lamp, or plaque. Sources are associated with specific organizations or institutions, so it is typically made up

of a group main designator, G, and a group sub-designator, g, both of which have the allowed values 0, ..., Z. Group codes are usually one of the following:

A for UA,
B for BSI,
C for CHORS,
D for SIO,
E for SRT,
G for GSFC,
H for SBRC,
I for NIST,
J for NRIFSF (Japan),
M for UM,
N for NOAA/MLML,
R for NRaD,
S for Satlantic,
U for UCSB, and
W for WFF.

Group sub-designators are separated into numbers and letters with numbers indicating spheres and plaques, and letters indicating lamps. Once combined, the common Gg codes are as follows:

A1 for the UA 10 in Spectralon plaque,
AA for FEL lamp F296,
AB for FEL lamp F297,

B1 for the BSI 10 in Spectralon plaque,
B2 for the BSI 20 in sphere,
B3 for the BSI 18 in Spectralon plaque,
BA for FEL lamp F310,
BB for FEL lamp F321,
BC for FEL lamp 91453,
BD for FEL lamp 91537,

C1 for the CHORS 40 in sphere with 150 W lamps,
C2 for the CHORS 40 in sphere with 275 W lamps,
C3 for the CHORS 18 in Spectralon plaque,
CA for FEL lamp 90572,
CB for FEL lamp 91348,
CC for FEL lamp 91534,
CD for FEL lamp 91533,
CE for FEL lamp 91349,

DA for FEL lamp F12L,

G1 for the GSFC 42 in sphere,
GA for FEL lamp F268,
GB for FEL lamp F269,
GC for FEL lamp F227,
GD for FEL lamp F315,
GE for FEL lamp F316,

H1 for the SBRC 20 in sphere,
HA for DXW lamp 19,
HB for DXW lamp SS27,
HW for the SBRC SeaWiFS instrument,

JA for FEL lamp F123,

MA for FEL lamp F12G,
MB for FEL lamp F12H,

N1 for the NOAA 10 in plaque (S/N OO1900),
N2 for the NOAA Gamma RS-10A (S/N HL1315),
N3 for the NOAA Gamma RS-10A (S/N HL382),
N4 for the NOAA Optronics 420M (S/N 92106057),
N5 for the NOAA Gamma RS-13 (S/N HL411),
NA for FEL lamp F307,
NB for FEL lamp F308,

RA for FEL lamp F265,

S1 for the Satlantic 18 in Spectralon plaque,
SA for DXW lamp S724,
SB for DXW lamp S721,
SC for FEL lamp 91604,
SD for FEL lamp 91591,
SE for FEL lamp 91615,

U1 for the UCSB 20 in sphere,
U2 for the UCSB 18 in Spectralon plaque,
U3 for the UCSB 10 in Spectralon plaque,
UA for FEL lamp F219,
UB for FEL lamp F393,
UC for FEL lamp F303,
UD for FEL lamp F304,
UE for FEL lamp F305,

W1 for the WFF 30 in sphere.

Note, the HW source identifier code is mentioned here, but it is reserved for the SBRC database as described in Chapter 5.

4.2.3 Instrument Identifier Code

The instrument identifier code, I, is a one-character code which identifies the instrument used to make the measurement. It has the allowed values 0, ..., Z, but is generally one of the following:

1 for the NIST absolute radiometer number 1,
2 for the NIST absolute radiometer number 2,
8 for the Satlantic 7-channel radiometer (S/N 8),
A for the UA (EOS) 7-channel transfer radiometer,
B for the BSI 6-channel transfer radiometer,
C for the CHORS 11-channel transfer radiometer,
I for the GSFC Optronics 746_ISIC,
P for the NIST PR714 radiometer,
S for the Satlantic 7-channel radiometer (S/N 26),
T for the GSFC Optronics 746_T, and
X for the SeaWiFS 6-channel transfer radiometer (SXR).

4.2.4 Configuration Code

The configuration code, CC, is a two-character code which identifies the configuration of the source and instrument during the measurement process. CC has the allowed values 0, ..., Z for each character position. Unknown or undetermined configurations are coded as __. The configuration codes for the various sources are as follows:

UA 10 in Spectralon Plaque A1

11 viewed at approximately 45° and ~150 cm.

BSI 10 in Spectralon Plaque B1

11 viewed at approximately 45° and ~150 cm.

BSI 20 in Sphere B2

00 aperture closed,
01 aperture in position 1,
02 aperture in position 2,
    ⋮
06 aperture in position 6.

BSI 18 in Spectralon Plaque B3

11 viewed at approximately 45° and ~150 cm.

CHORS Spheres C1 *and* C2

00 0.000–0.000–0.000–0.000 (all apertures closed),
01 1.250–1.250–1.250–1.250,
02 0.875–0.875–1.080–1.080,
03 0.628–0.628–0.766–0.766,
04 0.438–0.438–0.531–0.531,
05 0.312–0.312–0.438–0.438,
06 0.000–0.312–0.000–0.438, and
07 0.312–0.000–0.438–0.000.

CHORS 18 in Spectralon Plaque C3

11 viewed at approximately 45° and ~150 cm.

GSFC Sphere G1

00 all lamps off,
01 1 lamp on,
02 2 lamps on,
    ⋮
16 16 lamps on.

SBRC Sphere H1

AA 6 5 W lamps and 2 4 W lamps on,
AB 1 45 W lamp on,
AC 1 45 and 1 200 W lamp on,
AD 6 5, 2 45 and 1 200 W lamp on,
00 all lamps off,
01 1 200 W lamp on,
02 2 200 W lamps on,
    ⋮
11 11 200 W lamps on.

Lamp Calibrations

C1 CHORS baffles with the lamp at ~100 cm,
C5 CHORS baffles with the lamp at ~50 cm,
G5 GSFC baffles with the lamp at ~50 cm, and
I5 NIST baffles with the lamp at ~50 cm.

NOAA 10 in Spectralon Plaque N1

11 viewed at approximately 45° and ~150 cm.

NOAA Gamma RS-10A (S/N HL1315) N2

WO without diffuser attachment, and
DA with diffuser attachment RS-70-1.

NOAA Gamma RS-10A (S/N HL382) N3

WO without diffuser attachment, and
DA with diffuser attachment RS-70-1.

NOAA Optronics 420M N4

01 slide=3, wheel=5, and distance ~100, and
02 slide=3, wheel=6, and distance ~40.

NOAA Gamma RS-13 (S/N HL411) N5

A2 aperture wheel=200,
A3 aperture wheel=300,
A4 aperture wheel=400, and
A9 aperture wheel=900.

Satlantic 18 in Spectralon Plaque S1

11 viewed at approximately 45° and ~150 cm.

UCSB 20 in Sphere U1

UA FEL lamp F219 used at approximately 50 cm and
UB FEL lamp F393 used at approximately 50 cm. (If a lamp other than a UCSB lamp is used, its two-character code is entered instead.)

UCSB 18 in Spectralon Plaque U2

11 viewed at approximately 45° and ~150 cm.

UCSB 10 in Spectralon Plaque U3

11 viewed at approximately 45° and ~150 cm.

WFF 30 in Sphere W1

00 apertures 1 and 2 open (no aperture cards inserted),
01 aperture 1 open and aperture 2 in position 1,
    ⋮
06 aperture 1 open and aperture 2 in position 6,
10 aperture 1 in position 1 and aperture 2 open,
11 aperture 1 in position 1 and aperture 2 in position 1,
    ⋮
66 aperture 1 in position 6 and aperture 2 in position 6.

4.2.5 Data Type

The data type, d, is a one-character code which identifies the type of data collected during the measurement process and has the allowed values A, ..., Z, but is generally one of the following:

A for averaged voltages,
C for (responsivity) calibration file,
D for derived file,
E for irradiance,
H for historical file,
L for radiance,
T for (text) documentation file,
V for voltage, and
_ for undetermined or unknown.

4.2.6 Run Type

The run type, r, is a one-character code which identifies the type of run executed during the measurement process and has the allowed values 0, ..., Z, but is generally one of the following:

C for sphere radiance calibration using a cosine collector,

I for sphere irradiance,

L for lamp irradiance calibration,

M for lamp to lamp ratios,

P for plaque radiance instrument calibration,

Q for sphere responsivity,

R for sphere radiance calibration using a calibrated radiometer,

S for sphere radiance instrument calibration,

X for system response file, and

_ for undetermined or unknown.

4.2.7 Sequence Character

The sequence character, s, identifies which run in a series of runs was executed during the measurement process and has the allowed values 0, ..., Z; the underscore (_) is reserved to denote inapplicable, undetermined, or unknown entries.

4.3 EXPANDED CONVENTION

The expanded file name uses the 0, ..., Z character set of the condensed file names and adds one more: the hyphen (-) which is used as a sub-field embedded separator. The underscore (_) is used for undetermined or unknown elements and the period (.) is used to separate the expanded name into smaller elements. Every expanded file name uses 22 characters starting with 1) a seven-character time stamp that identifies when the data was taken, followed by 2) a six-character primary test code identifying what was measured, then 3) a three-character code giving the instrumentation used and its configuration during data collection, and finally 4) a three-character sequence identifying the type of data. The expanded file naming convention is as follows:

YYMMMDD.GEEEEE.ICC.drs,

where, YYMMMDD is a Gregorian time stamp (YY is the decadal year with: 91 representing 1991, 92 representing 1992, 00 representing 2000, and 25 representing 2025, MMM is an abbreviated month name which uses the first three characters of the month, and DD is the day of the month); G is the group code (described earlier in the *Source Identifier Code* section of the condensed file naming convention); EEEEE is an expanded source identifier described below; I is the *Instrument Identifier Code*; CC is the *Configuration Code*; d is the *Data Type*; r is the *Run Type*; and s is the *Sequence Character*. Note that YY covers only the range 1991–2025, for consistency with the condensed SIRREX

convention described in Section 4.2. Purposely there is no definitive file extension on the expanded file name, so the user can append one based on local requirements.

The complete expanded source identification code is made up of the group code G and an expanded source name EEEEE. The hyphen is used to ensure a dense five-character source name. The common GEEEEE codes are as follows:

A-10in for UA 10 in Spectralon plaque,

A-F296 for FEL lamp F296,

A-F297 for FEL lamp F297,

B-10in for BSI 10 in Spectralon plaque,

B-18in for BSI 18 in sphere,

B-20in for BSI 20 in sphere,

B-F310 for FEL lamp F310,

B-F321 for FEL lamp F321,

B91453 for FEL lamp 91453,

B91537 for FEL lamp 91537,

C-18in for CHORS 18 in Spectralon plaque,

C40in1 for CHORS 40 in sphere w/150 W lamps,

C40in2 for CHORS 40 in sphere w/275 W lamps,

C90572 for FEL lamp 90572,

C91348 for FEL lamp 91348,

C91349 for FEL lamp 91349,

C91533 for FEL lamp 91533,

C91534 for FEL lamp 91534,

D-F12L for FEL lamp F12L,

G-42in for GSFC 42 in sphere,

G-F227 for FEL lamp F227,

G-F268 for FEL lamp F268,

G-F269 for FEL lamp F269,

G-F315 for FEL lamp F315,

G-F316 for FEL lamp F316,

H-20in for SBRC 20 in sphere,

H-SS27 for DXW lamp SS27,

HDXW19 for DXW lamp 19,

HSWiFS for SeaWiFS instrument,

J-F123 for FEL lamp F123,

M-F12G for FEL lamp F12G,

M-F12H for FEL lamp F12H,

N-10in for NOAA 10 in plaque (S/N 001900),

N-420M for NOAA Optronics 420M (S/N 92106057),

N-F307 for FEL lamp F307,

N-F308 for FEL lamp F308,

NG1315 for NOAA Gamma RS-10A (S/N HL1315),

N-G382 for NOAA Gamma RS-10A (S/N HL382),

N-G411 for NOAA Gamma RS-13 (S/N HL411),

R-F265 for FEL lamp F265,

S-18in for Satlantic 18 in Spectralon plaque,

S-S721 for DXW lamp S721,

S-S724 for DXW lamp S724,

S91591 for FEL lamp 91591,

S91604 for FEL lamp 91604,

S91615 for FEL lamp 91615,

```
U-10in  for UCSB 10 in sphere,
U-18in  for UCSB 18 in sphere,
U-20in  for UCSB 20 in sphere,
U-F219  for FEL lamp F219,
U-F303  for FEL lamp F303,
U-F304  for FEL lamp F304,
U-F305  for FEL lamp F305,
U-F393  for FEL lamp F393, and
W-30in  for WFF 30 in sphere.
```

Note again, the HSWiFS source identifier code is mentioned here, but it is reserved for the SBRC database as described in Chapter 5.

4.4 ARCHIVING PROCEDURES

The SIRREX protocol specifies a header format which is inserted at the beginning of every file before it is archived. Any relevant information in an existing header is incorporated into the SIRREX keyword vocabulary. Additional notations are incorporated as comment lines, which begin with an exclamation point (!). The requirement for tab-delimited text and a sequence column often requires modification of the original data. While the data is being modified, a consistent numeric format is used for each column of data. This format is chosen based on the range and precision of the data taken by a particular instrument. Using a consistent numeric format allows a quick visual inspection of the data to locate errors.

Investigators participating in a SIRREX are encouraged to provide data in a form as close to the established protocol as is feasible. Many radiometers brought to SIRREXs use commercial software that creates output in a specific format, which mandates modifying the data before the file can be archived. Each instrument has its own log sheet which is used to record all the relevant information during each measurement. These log sheets are then used to generate the header and archive filename for each file.

4.4.1 Archived File Example

An example file, 36JG1P16.LS3, with a complete header block, but abbreviated data block, is presented below:

```
/begin_header
/aperture_diameter=16[in]
/color_temperature=2964[K]
/column_headers=[nm],[W_m^{-2}_sr^{-1}_nm^{-1}]
/comment=Small_Field_Stop
/configuration=16_Lamps_On
/data_type=Radiance
/date=1993JUN19
/experiment=SIRREX_2
/group_identifier=GSFC
/group_sub_identifier=42_Inch_Sphere
/information_file=714_doc.t_2
/instrument_identifier=NIST_PR714
```

```
/investigators=Unknown
/lab_temperature=27[C]
/lamp_current=bank1_6.502[A],bank2_6.499[A]
/number_data_rows=173
/original_file_name=36IGSP00.VR3
/pr714_field_stop=0.125_x_0.125[deg]
/pr714_integration_time=628[msec]
/pr714_lens_distance=1.3[m]
/run_type=Sphere_Radiance_Instrument_Calibration
/sequence_number=3
/source_distance=51.75[in]
/sphere_position=centered_(0,0)
/time=17:10:03[PDT]
/total_radiance=207.1[W_m^{-2}_sr^{-1}]
/wavelength_range=380>1068;4[nm]
/wavelength_resolution=5[nm]
!Sphere was turned on at 15:49.
!PR714 was turned on at 9:30.
/end_header@
380 1.833e-002
384 1.962e-002
388 2.131e-002
     .
     .
1060 3.963e-001
1064 3.988e-001
1068 4.021e-001
```

4.4.2 Condensed File Name Example

The sample file name 36JG1P16.LS3 represents the following:

Year 3 = 1993,
Month 6 = JUN,
Day J = 19,
Source Identifier G1 = GSFC 42 in sphere,
Instrument Identifier P = NIST PR714 radiometer,
Configuration Code 16 = 16 lamps on,
Data Type L = Radiance,
Run Type S = Sphere radiance instrument calibration, and
Sequence Number 3 = Run number three.

It is worth noting again that a verbose description of the file name is included in the keywords:

```
/configuration,
/data_type,
/date,
/instrument_identifier,
/run_type,
/sequence_number, and
/source_identifier.
```

The configuration code used depends on the source used. A configuration code of 16 indicates 16 lamps were used if the source was the GSFC 42 in sphere.

4.4.3 Expanded File Name Example

Using the same sample file, the date stamp of 13 June 1993 would be coded as 93JUN13. The GSFC 42 in sphere is represented as G-42in. Instrument and configuration codes remain unchanged, (P16 in the example file), as do data type, run type, and sequence codes, e.g., LS3. The expanded file name is coded as:

93JUN13.G-42in.P16.LS3.

The expanded format uses only 22 of the 31 characters normally available to many operating systems capable of extended file naming. The remaining characters are available for use in local applications, perhaps for tracking post-processing of the data.

4.5 KEYWORD GLOSSARY

The keyword glossary is composed of two sets: required and approved keywords. The former must appear in every SIRREX data file, while the latter appear only as needed to fully describe the contents of the file. Of course, local use of a data file permits the addition of locally defined keywords, perhaps for tracking post-processing of the data.

4.5.1 Required Keywords

The required keywords for the SIRREX data sets are as follows:

/begin_header This keyword does not have a value and must be the first entry in the series of lines comprising the header block.

/column_headers=$\langle s \rangle$ This is a comma-delimited list of the headers for the columnar data in the file. Usually the ith entry in the list is associated with the ith column. In cases where a sequence of data is taken, the data columns are indicated by a range. Bands 1 through 6 would be band1>band6, and 25 data points and their average would be represented as raw01>raw25,average.

/configuration=$\langle s \rangle$ This string is a verbose description of the specific configuration used as part of the file name and describes the physical configuration or setup used for the measurement in question. Usually orientation, distances, and other parameters are specific to a type of measurement.

/data_type=$\langle s \rangle$ This is a verbose version of the data type given in the file name.

/date=$yyyymmdd$ In this format, $yyyy$ is the year, mmm is the abbreviated three-character Gregorian month of the year, and dd is the day of the Gregorian month. The month abbreviations use the first three characters of the month: Jan is January, Feb is February, etc.

/end_header@ This keyword does not have a value and must be the last entry in the series of lines comprising the header block. Note the at-sign (@) at the end.

/experiment=$\langle s \rangle$ This entry identifies the experiment associated with the data collected, e.g., SIRREX_2.

/group_identifier=$\langle s \rangle$ This string identifies the group conducting the test, e.g., CHORS.

/group_sub_identifier=$\langle s \rangle$ This string identifies the specific lamp, sphere, or plaque used during the test.

/instrument_identifier=$\langle s \rangle$ This string identifies the instrument used to conduct the test.

/investigators=$\langle s \rangle$ The string for this keyword contains the names of the individuals (separated by commas) who collected the data contained in the file.

/number_data_rows=$\langle n \rangle$ This is the number of data rows (or lines) after the /end_header@ keyword. If this number is different than the actual number of data rows in the file, there has been some kind of file corruption.

/original_file_name=$\langle s \rangle$ Most data is acquired on computer systems that have already adopted a particular file naming convention. Once the data from such a device is transferred to the SIRREX database, it receives a new name based on the condensed SIRREX naming convention. This entry is reserved for keeping track of the original file name.

/run_type=$\langle s \rangle$ This string is a verbose version of the run type recorded in the file name and identifies the type of run executed during the measurement process. It has the allowed values given in the *Run Type* section discussed earlier.

/sequence_number=$\langle n \rangle$ This number indicates which experiments, in a series of experiments, are contained within this file. In general, the sequence numbering is also used to distinguish archive file numbers that would otherwise be identical.

There is no specific requirement regarding the ordering of the keywords in the files, other than the /begin_header and /end_header@ keywords, which must be first and last respectively. It is recommended, however, that the second through next to last keywords be kept alphabetical to maximize the readability of the file.

4.5.2 Approved Keywords

The approved keywords for the SIRREX data sets are as follows:

/aperture_diameter=$\langle n \rangle [u]$ This is the diameter of the source aperture.

/BSI_channels=$\langle n \rangle$ These are the channels used on the BSI radiometer, nominally 1>6.

/BSI_units=⟨s⟩ These are the units of the BSI radiometric measurement, nominally volts.

/calibration_file=⟨s⟩ This is the file containing the calibration information for the radiometer in question. The file contains the needed information to convert from raw data to radiance.

/color_temperature=⟨n⟩[u] This field indicates the equivalent blackbody temperature of the source in degrees Kelvin [K].

/comment=⟨s⟩ This field is intended for comment entries included on most of the data logs.

/entrance_slit=⟨n⟩[u] This keyword is the monochrometer entrance slit width.

/exit_slit=⟨n⟩[u] This is the width of the monochrometer exit slit.

/gain=⟨n⟩ This is the viewing instrument gain and should be recorded as a complete factor rather than just an exponent, for example, 10^{2} or 100 rather than 2.

/information_file=⟨s⟩ This is a text (ASCII) file containing instrument and file format information.

/internal_averages_per_reading=⟨n⟩ This is the number of internal averages per reading for the BSI radiometer, nominally 100.

/lab_temperature=⟨n⟩[u] The temperature of the laboratory where the data was collected in units of Fahrenheit [F] or Celsius [C].

/lamp_current=⟨n⟩[u] The current applied to the lamp(s) used. The units are usually amps [A]. A comma-delimited list can be used for multiple lamp scenarios, in which case, values for particular banks are indicated by preceding the value with bank1_, bank2_, etc., indicators. For example, /lamp_current=bank1_10[A], bank2_20[A].

/lamp_voltage=⟨n⟩[u] The voltage applied to the lamp(s) used. The units are usually volts [V]. A comma-delimited list can be used for multiple lamp scenarios, in which case, values for particular banks are indicated by preceding the value with bank1_, bank2_, etc., indicators. For example, /lamp_voltage=bank1_10[V], bank2_20[V].

/number_of_readings_per_external_average=⟨n⟩ This is the number of BSI readings, where each reading is a number of internal averages, which are averaged externally, typically 30.

/plaque_distance=⟨n⟩[u] This is the distance between the lamp and the plaque.

/plaque_viewing_angle⟨n⟩[u] This keyword is the angle of the viewing radiometer with respect to the plaque, nominally 45[deg].

/pr714_field_stop=⟨s⟩ These are the field stops used with the PR714 radiometer, either 0.125_x_0.125[deg] or 0.5_x_1.5[deg].

/pr714_integration_time=⟨n⟩[u] This is the integration time of the PR714.

/pr714_lens_distance=⟨n⟩[u] The focal distance of the lens mounted on the PR714, as determined by the scale imprinted on the lens.

/radius_one=⟨n⟩[u] This keyword gives the aperture radius of the source.

/radius_two=⟨n⟩[u] This is the aperture radius of the collector.

/reference_lamp_id=⟨s⟩ This keyword is the conventional name for the reference lamp used, like, F269, 91453, etc., used to illuminate the sphere or plaque.

/relative_humidity=⟨n⟩[u] This is the relative humidity of the laboratory air during the test. The units are usually [%].

/serial_number=⟨s⟩ This is the serial number of the recording instrument. It is shown here as a general character string, since many serial numbers include letters.

/source_distance=⟨n⟩[u] This is the distance between the viewing radiometer and the source.

/sphere_diameter=⟨n⟩[u] This keyword is the diameter of the source sphere.

/sphere_position=⟨s⟩ This is the viewing position of the radiometer with respect to the source aperture, nominally centered_(0,0).

/sxr_channels=⟨n⟩ This entry gives the channels used on the SXR, nominally 1>6.

/sxr_lens=⟨s⟩[u] This entry gives the f-stop and distance used with the SXR radiometer; nominally, this keyword has a value of f/1.4_1.2[m].

/sxr_temperature=⟨n⟩[u] This is the temperature setting of the thermoelectric cooler for the SXR.

/time=⟨s⟩[u] The format for ⟨s⟩ is hh:mm:ss.ss, where hh is the 24-hour hour of the day, mm is the minute of the hour, and ss.ss is the decimal second of the hour. The ss.ss field is optional and the [u] field is used to denote GMT, EST, EDT, etc., nominally [GMT].

/total_radiance=⟨n⟩[u] This is the total radiance collected in units of W_m^{-2}_sr^{-1}.

/ua_aperture=⟨n⟩[u] The aperture setting on the UA radiometer.

/ua_channels=⟨s⟩ The entry gives channels used with the UA radiometer.

/ua_temperature=$\langle n\rangle$[u] The temperature of the UA radiometer.

/wavelength_range=$\langle n\rangle$[u] This is the wavelength range for the data expressed as $n_1 > n_2;n_3$, where n_1 is the lowest value in the range, n_2 is the largest value in the range, and n_3 is the increment value. The units are usually [nm].

/wavelength_resolution=$\langle n\rangle$[u] This is the spectral resolution of the instrument performing the measurement. Units are typically [nm].

4.6 DISCUSSION

The SIRREX archive contains files stored in an ASCII text format that is independent of the platform in use.

Each file has a header block containing information regarding when, where, how, and by whom the data was taken. This header also contains information specifying how many rows and columns of data to expect. Headers from original data files are placed into SIRREX keywords, when appropriate, as well as information from the log sheet. A sequence column is added if necessary. Each column of data is written in a single numeric format before the file is placed in the archive. The result is a file that is usable on many platforms and has descriptive header information in a known format that is easily parsed for automated processing. This header contains the details of how the data was obtained and how much and what type of data is contained in the file.

Chapter 5

The SBRC Database

TODD L. WESTPHAL
General Sciences Corporation, Laurel, Maryland

YUNTAO GE
University Space Research Association, Columbia, Maryland

STANFORD B. HOOKER
NASA Goddard Space Flight Center, Greenbelt, Maryland

ABSTRACT

This document describes the naming convention and procedures used in archiving the SeaWiFS prelaunch calibration data obtained from SBRC. The archive contains data from BTR, ASR, LSF, and temperature dependence measurements, as well as images of the sun, moon, sky, and an integrating sphere. This data set contains characterization information on the SeaWiFS instrument provided by the manufacturer. As such, it serves as a baseline to the operation of the instrument on orbit.

5.1 INTRODUCTION

The SBRC database is a subset of the SIRREX database and closely follows that architecture as described in Chapter 4. As part of SeaBASS, the SBRC archive has both a condensed *and* an expanded file naming convention. The former uses the so-called *eight-point-three* 11-character format used by MS-DOS computers; the latter is intended for use on computers that support a 31-character file name. All files are archived using the condensed file name, but users can request files in the expanded naming convention.

5.2 CONDENSED CONVENTION

Every condensed file name starts with a three-character time stamp that identifies when the data was taken, followed by a five-character primary test code identifying what was measured and with what instrumentation, and then a three-character extension identifying the type of data recorded. The condensed file naming convention is as follows:

DDDPPPPP.eee

where DDD is the date stamp, PPPPP is the primary test code, and eee is the file extension. The date stamp is composed of three characters: YMD where, Y is the year code, M is the month code, and D is the day of the month code. The primary test code is composed of two sub-codes: SSRRR where, SS is a two-character source identifier, and RRR is a three-character run type. The file extension is divided into three codes: dbs where, d is a one-character data type, b is a one-character band indentifier, and s is a one-character sequential run character (0,...,9). Thus, the complete condensed file name convention is:

DDDSSRRR.dbs

where, once again, the condensed convention makes full use of the *eight-point-three* 11-character format designed for MS-DOS PCs.

5.2.1 Condensed Date Code

The condensed time code is a three-character code, YMD, and is decoded as follows: Y has the allowed values 1,...,Z, where 1=1991, 2=1992, 3=1993, A=2000, and Z=2025; M has the allowed values 1,2,3,...,A,B,C, where 1=Jan., 2=Feb., 3=Mar., A=Oct., B=Nov., and C=Dec.; and D has the allowed values 1,2,3,...,A,B,...,V, where 1=1, 2=2, 3=3, A=10, B=11, and V=31. If any part of the date field is not known, the underscore (_) is used for the unknown component.

5.2.2 Source Identifier Code

The source identifier code is a two-character code which identifies the data source as SBRC and the SeaWiFS instrument. It is always equal to HW.

5.2.3 Condensed Run Type

The run type has different allowed values depending on the data type (d), although all data types allow DOC to indi-

cate their information file. All DOC files use the associated data type as the value for this keyword. For example, the information file for image data contains /run_type=Image.

For a data type of A, all files are assigned the run type DAT.

For a data type of B, the first two digits indicate the width of the bright target in pixels, and the third character represents the illumination level. The allowed run types for B data are:

01M for a 1 pixel wide target, mid-level illumination,

01H for a 1 pixel wide target, high-level illumination,

03M for a 3 pixel wide target, mid-level illumination,

03H for a 3 pixel wide target, high-level illumination,

06H for a 6 pixel wide target, high-level illumination,

10M for a 10 pixel wide target, mid-level illumination, and

10H for a 10 pixel wide target, high-level illumination,

For a data type of D, all files are assigned the run type LIN. For a data type of I, the three-character code describes what the sensor was viewing and has the following allowed values:

INT for an image of the integrating sphere,

LUN for a lunar image,

SKY for an image of the sky, and

SOL for a solar image.

For a data type of L, the three-character code describes in which direction the line spread function was measured and has the following allowed values:

ASC for an along-scan measurement, and

ATR for an along-track measurement,

For a data type of T, the file is assigned the run type FCT.

5.2.4 Condensed Data Type

This one-character code establishes the type of data collected and has the following allowed values:

A for ASR,

B for BTR,

D for linearity data,

I for images,

L for LSF, and

T for temperature dependence and telemetry coefficients.

5.2.5 Band Code

The band code is comprised of a single character and has the allowed values of 1, ..., 8, A and N. A number indicates which one of the possible eight bands is in the data file; whereas, A indicates all eight bands are included and N indicates no bands are included. The N designation is used only on documentation files; all other files must contain data from at least one band.

5.2.6 Sequence Number

This digit represents which particular measurement in a series of similar measurements is present in the data file. The allowed values are 0, ..., 9.

5.3 EXPANDED CONVENTION

The expanded file name uses the 0, ..., Z character set of the condensed file names. The underscore (_) is used for undetermined or unknown elements and the period (.) is used to separate the expanded name into smaller elements. Every expanded file name uses 22 characters starting with 1) a seven-character time stamp that identifies when the data was taken, followed by 2) a six-character primary test code identifying what was measured, then 3) a three-character code giving the instrumentation used and its configuration during data collection, and finally 4) a three-character sequence identifying the type of data. The expanded file naming convention is as follows:

YYMMMDD.EEEEEE.RRR.dbs,

where, YYMMMDD is a Gregorian time stamp (YY is the decadal year with 50 representing 1950 and 49 representing 2049, MMM is an abbreviated month name which uses the first three characters of the month, and DD is the day of the month); EEEEEE is an expanded source identifier which is always set to HSWiFS for the SeaWiFS instrument; RRR is the *Run Type*; d is the *Data Type*; b is the *Band Code*; and s is the *Sequence Character*. Purposely, there is no definitive file extension on the expanded file name, so the user can append one based on local requirements.

5.4 ARCHIVING PROCEDURES

This protocol specifies a header format which is added at the beginning of every archived file, as for all SeaBASS files. All relevant information in an existing header is incorporated into the SBRC keyword vocabulary. The requirement for tab-delimited data columns and a sequence column often require modifying the original data. In cases where there was a pixel number or wavelength column, this was used in place of a sequence number. Although neither of these columns starts with 1, both are integer values that are incremented by one.

While the data is being modified, a consistent numeric format is used for each column of data. This format is chosen based on the range and precision of the data. Using a consistent numeric format allows a quick visual inspection of the data to locate errors.

5.4.1 Archived File Example

This example is identical in form to an actual file in the SBRC archive. Some of the keyword values have been edited to fit in the space provided. Also, not all 1288 lines of data are presented.

Sample Data File: 383HWSKY.IA5

```
/begin_header
/band=All
/column_headers=Pixel_#,Band_1>Band_8
/comment=looking_due_east,34.5_deg_elevation
/data_type=Image
/date=1993MAR07
/gain=B1_2,B2_2,B3_2,B4_2,B5_2,B6_2,B7_2,B8_2
/information_file=45AHWDOC.INO
/instrument_identifier=SeaWiFS
/investigators=Alan_Holmes
/mirror_side=Side_A_and_Side_B
/modification_status=Before_Stray_Light_Mods
/number_data_rows=1288
/original_file_name=sky1145
/run_type=Sky
/sequence_number=6
/time=11:45[PST]
/end_header@
-001 1023 0000 1023 0000 1023 0000 1023 1023
0000 0022 0018 0019 0021 0017 0016 0020 0020
0001 1023 1023 1023 1023 1023 1023 1023 1023
0002 1023 1023 1023 1023 1023 1023 1023 1023
0003 1023 1023 1023 1023 1023 1023 1023 1023
0004 1023 1023 1023 1023 1023 1023 1023 1023
0005 1023 1023 1023 1023 1023 1023 1023 1023
 .
 .
 .

1280 1023 1023 1023 1023 1023 1023 1023 1023
1281 1023 1023 1023 1023 1023 1023 1023 1023
1282 1023 1023 1023 1023 1023 1023 1023 1023
1283 1023 1023 1023 1023 1023 1023 1023 1023
1284 1023 1023 1023 1023 1023 1023 1023 1023
1285 1023 1023 1023 1023 1023 1023 1023 1023
1286 1023 1023 0000 0000 1023 1023 0000 1023
```

5.4.2 Condensed File Name Example

The sample file name, 383HWSKY.IA5, represents the following:

Year 3 = 1993
Month 8 = August
Day 3 = 3
Source Identifier HW = SBRC and the SeaWiFS
 instrument
Run Type SKY = Sky
Data Type I = Image
Band Number A = All Bands
Sequence Number 5 = 5th run

Note that a verbose description of the file name is included in the keywords: /date, /instrument, /run_type, /data_type, /band, and /sequence_number.

5.4.3 Expanded File Name Example

Using the same file name, 383HWSKY.IA5, the expanded codes are:

Year 1993 = 93
Month August = AUG
Day 3 = 03
Source Identifier HW = SBRC and the SeaWiFS
 instrument
Run Type Sky = SKY
Data Type Image = I
Band Number All Bands = A
Sequence Number 5 = 5

So the expanded file name is:

93AUG03.HSWiFS.SKY.IA5

This format uses only 22 of the 31 characters available, leaving the remaining characters available for use in post-processing applications.

5.5 KEYWORD GLOSSARY

The keyword glossary consists of required and approved keywords. Required keywords will be found in all archived files, while approved keywords appear whenever the relevant value is known.

5.5.1 Required Keywords

The required keywords for the SBRC database follow in alphabetical order:

/band=$\langle s \rangle$ This string indicates the band measured and has the allowed values of 1>8, A and N.

/begin_header This keyword does not have a value and must be the first entry in the series of lines comprising the header block.

/column_headers=$\langle s \rangle$ This keyword is a descriptive list of headers, which is comma-delimited, for the columnar data in the file. Usually the ith entry in the list is associated with the ith column. In cases where a sequence of data is taken, the data columns are indicated by a range. For example, bands 1 through 6 would be recorded as band_1>band_6 and the scan number followed by pixels 865 through 1285 would be represented as Scan_Number,pixel_865>pixel_1285. Each column header is a description of the data, and no abbreviations are used.

/data_type=$\langle s \rangle$ This is a verbose description of the data type recorded in the file name. See Section 5.2.4 for allowed values.

/date=$yyyymmmdd$ In this format, $yyyy$ is the year, mmm is the abbreviated three-character Gregorian month of the year, and dd is the day of the Gregorian month. The month abbreviations use the first three characters

33

of the month: JAN is January, FEB is February, etc. In cases where the day is unknown, it is not used (for example, 1994DEC).

/end_header@ This keyword does not have a value and must be the last entry in the series of lines comprising the header block. Note the at-sign (@) at the end.

/gain=⟨s⟩ This indicates the gain setting for each band. Each gain value is preceded by a band designation. For example, band 1 at gain 1 and band 3 at gain 2 would be designated B1_1,B3_2.

/instrument_identifier=⟨s⟩ This string indentifies the instrument used during the test. For SBRC data, the instrument is always SeaWiFS.

/investigators=⟨s⟩ The string for this keyword contains the names of the individuals (separated by commas) who collected the data contained in the file.

/modification_status=⟨s⟩ The string for this keyword indicates whether the data was taken before or after the stray light modifications were made to the Sea-WiFS instrument.

/number_data_rows=⟨n⟩ This is the number of data rows (or lines) after the /end_header@ keyword. If this number is different than the actual number of data rows in the file, there has been some kind of file corruption.

/original_file_name=⟨s⟩ This string provides a record of the file name prior to archiving. The original files are held in a separate archive complete with the original format and headers. The original files provide a baseline reference to the archived data.

/run_type=⟨s⟩ This is a verbose description of the run type recorded in the file name. Refer to Section 5.2.3 for allowed values.

/sequence_number=⟨n⟩ This is identical to the sequence number recorded in the file name.

5.5.2 Approved Keywords

The approved keyword list for the SBRC data set is:

/collection_aperture=⟨n⟩[u] This describes the aperture setting used during an ASR measurement.

/comment=⟨s⟩ This string is a record of any comment entered in the original header during testing.

/counts_per_amp=⟨n⟩[u] This specifies the value used to convert digital counts into amperes during an ASR measurement.

/factor=⟨n⟩ This is a unitless number used to force this computed data to agree with data obtained from a sphere measurement. A number greater than one implies more amps than expected.

/filter=⟨n⟩[u] This specifies the type and thickness of the Schott color glass filter used.

/ifov=⟨n⟩[u] This specifies the instantaneous field-of-view of the SeaWiFS instrument during an ASR measurement.

/information_file=⟨s⟩ This string provides the file name of an ASCII text file containing additional information regarding the measurement setup used and the format of the data obtained.

/lamp=⟨s⟩ This string indicates the number and wattage of lamps used during a BTR test.

/mirror_side=⟨s⟩ This string describes which side of the half-angle mirror was used during the test. If both sides of the mirror were used, it will be indicated here and there will be a column to indicate which side was measured for each data point.

/slit=⟨n⟩[u] This specifies the horizontal and vertical dimensions of the target slit in a BTR measurement.

/time=hh:mm:ss[u] This keyword records the time in the hours:minutes:seconds format. The time zone is indicated in the unit field as a three character abbreviation, EST for Eastern Standard Time, GMT for Greenwich Mean Time, etc.

/wavelength_range=⟨n⟩[u] This string indicates the spectral range, increment, and units used during an ASR measurement. For example, a measurement starting at 400 nm, and ending at 850 nm using a 4 nm increment would be represented as: 400>850;4[nm].

5.6 DISCUSSION

The SBRC archive contains files compliant with the SeaBASS protocol. Each file's header contains the original header information as well as information specific to the data type and archive file name. The original image files were split into separate bands to reduce the size of the files and eliminate embedded headers between bands. Line numbers were added to any file that did not already have them as a means of determining gaps in the data. Files that contained space-delimited data were converted to tab-delimited. The /column_headers keyword provides a list of what data appears in each column. The result is a self-contained file header that is similar to that of the SIRREX database.

GLOSSARY

ASCII American Standard Code for Information Interchange

ASR Absolute Spectral Response

BATS Bermuda Atlantic Time Series
BBOP Bermuda Bio-Optical Profiler
BOFS British Ocean Flux Study
BSI Biospherical Instruments, Incorporated
BTR Bright Target Recovery

CHORS Center for Hydro-Optics and Remote Sensing (San Diego State University)
COTS Commercial Off-The-Shelf
CVT Calibration and Validation Team
CZCS Coastal Zone Color Scanner

DAAC Distributed Active Archive Center
DAT Digital Audio Tape
DEC Digital Equipment Corporation
DXW Not an acronym, but a lamp designator.

E-mail Electronic Mail
EDT Eastern Daylight Time
EOS Earth Observing Satellite
EqPac Equatorial Pacific
EST Eastern Standard Time

FEL Not an acronym, but a lamp designator.
FTP File Transfer Protocol

GLOBEC Global Ocean Ecosystems dynamics
GMT Greenwich Mean Time
GSFC Goddard Space Flight Center
GUI Graphical User Interface

IBM International Business Machines
IDL Interactive Data Language
IP Internet Protocol
ISIC Integrating Sphere Irradiance Collector

JARE Japanese Antarctic Research Expedition
JGOFS Joint Global Ocean Flux Study

LDEO Lamont-Doherty Earth Observatory (Columbia University)
LOC Local Time
LSF Line Spread Function

MIT Massachusetts Institute of Technology
MLML Moss Landing Marine Laboratory
MO Magneto-Optical
MOBY Marine Optical Buoy
MOCE Marine Optical Characterization Experiment
MS-DOS Microsoft/Disk Operating System

NASA National Aeronautics and Space Administration
NET NIMBUS Experiment Team
NIMBUS Not an acronym, a series of NASA experimental satellites containing a wide variety of atmosphere, ice, and ocean sensors.
NIST National Institute of Standards and Technology
NOAA National Oceanic and Atmospheric Administration
NRaD Naval Research and Development
NRIFSF National Research Institute of Far Seas Fisheries (Japan)

OSC Orbital Sciences Corporation

PAR Photosynthetically Available Radiation
PC (IBM) Personal Computer
PR Photo Research

QC Quality Control

RDBMS Relational Database Management System

SBRC (Hughes) Santa Barbara Research Center
SeaBASS SeaWiFS Bio-Optical Archive and Storage System
SEAPAK Not an acronym, an image display and analysis package developed at GSFC.
SeaWiFS Sea-viewing Wide Field-of-view Sensor
SGI Silicon Graphics, Incorporated
SI International System of Units (SI), or *Système International d' Unitès*
SIO Scripps Institution of Oceanography
SIRREX SeaWiFS Intercalibration Round-Robin Experiment
S/N Serial Number
SRT Sigma Research Technology, Incorporated
STM Science Team Member
Sun Sun Microsystems
SXR SeaWiFS Transfer Radiometer

UA University of Arizona
UCSB University of California at Santa Barbara
UM University of Miami
UNIX Not an acronym, a computer operating system.

WFF Wallops Flight Facility (NASA)
WHOI Woods Hole Oceanographic Institute
WOCE World Ocean Circulation Experiment

REFERENCES

Barnes, R.A., A.W. Holmes, W.L. Barnes, W.E. Esaias, and C.R. McClain, 1994: The SeaWiFS Prelaunch Radiometeric Calibration and Spectral Characterization. *NASA Tech. Memo. 104566, Vol. 23*, S.B. Hooker and E.R. Firestone, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, (submitted).

Clark, D.K., 1981: Phytoplankton algorithms for the Nimbus-7 CZCS. *Oceanography from Space,* J.F.R. Gower, Ed., Plenum Press, 227–238.

Evans, R.H., and H.R. Gordon, 1994: CZCS "system calibration:" A retrospective examination, *J. Geophys. Res.,* **99**, 7,293–7,307.

Flierl, G., D. Glover, J. Bishop, and S. Paranjpe, 1993: *The JGOFS Distributed Object Oriented Data System User Guide.* Massachusetts Institute of Technology, Cambridge, Massachusetts, 90 pp.

Habermann, T., 1991: Freeform—A Flexible System of Format Specifications For Data Access. National Geophysical Data Center, NOAA, 37 pp.

Hooker, S.B., W.E. Esaias, G.C. Feldman, W.W. Gregg, and C.R. McClain, 1992: An Overview of SeaWiFS and Ocean Color. *NASA Tech. Memo. 104566, Vol. 1*, S.B. Hooker and E.R. Firestone, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, 24 pp., plus color plates.

——, W.E. Esaias, and L.A. Rexrode, 1993: Proceedings of the First SeaWiFS Science Team Meeting, *NASA Tech. Memo. 104566, Vol. 8*, S.B. Hooker and E.R. Firestone, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, 61 pp.

McClain, C.R., M. Darzi, J. Firestone, E-n. Yeh, G. Fu, and D. Endres, 1991a: SEAPAK Users Guide, Version 2.0, Vol. I—System Description. *NASA Tech. Mem. 100728*, NASA Goddard Space Flight Center, Greenbelt, Maryland, 158 pp.

——, ——, ——, ——, ——, and ——, 1991b: SEAPAK Users Guide, Version 2.0, Vol. II—Descriptions of Programs. *NASA Tech. Mem. 100728*, NASA Goddard Space Flight Center, Greenbelt, Maryland, 586 pp.

——, W.E. Esaias, W. Barnes, B. Guenther, D. Endres, S.B. Hooker, G. Mitchell, and R. Barnes, 1992: Calibration and Validation Plan for SeaWiFS. *NASA Tech. Memo. 104566, Vol. 3*, S.B. Hooker and E.R. Firestone, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, 41 pp.

Mitchell, B.G., and O. Holm-Hansen, 1991: Bio-optical properties of Antarctic Peninsula waters: differentiation from temperate ocean models. *Deep-Sea Res.*, **39**, 1,009–1,028.

Mueller, J.L., 1993: The First SeaWiFS Intercalibration Round-Robin Experiment, SIRREX-1, July 1992. *NASA Tech. Memo. 104566, Vol. 14*, S.B. Hooker and E.R. Firestone, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, 60 pp.

——, and R.W. Austin, 1992: Ocean Optics Protocols for SeaWiFS Validation. *NASA Tech. Memo. 104566, Vol. 5*, S.B. Hooker and E.R. Firestone, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, 43 pp.

——, B.C. Johnson, C.L. Cromer, J.W. Cooper, J.T. McLean, S.B. Hooker, and T.L. Westphal, 1994: The Second SeaWiFS Intercalibration Round-Robin Experiment, SIRREX-2, June 1993. *NASA Tech. Memo. 104566, Vol. 16*, S.B. Hooker and E.R. Firestone, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, 121 pp.

Woodward, R.H., R.A. Barnes, C.R. McClain, W.E. Esaias, W.L. Barnes, and A.T. Mecherikunnel, 1993: Modeling of the SeaWiFS Solar and Lunar Observations. *NASA Tech. Memo. 104566, Vol. 10*, S.B. Hooker and E.R. Firestone, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, 26 pp.

THE SeaWiFS Technical Report Series

Vol. 1

Hooker, S.B., W.E. Esaias, G.C. Feldman, W.W. Gregg, and C.R. McClain, 1992: An Overview of SeaWiFS and Ocean Color. *NASA Tech. Memo. 104566, Vol. 1*, S.B. Hooker and E.R. Firestone, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, 24 pp., plus color plates.

Vol. 2

Gregg, W.W., 1992: Analysis of Orbit Selection for SeaWiFS: Ascending vs. Descending Node. *NASA Tech. Memo. 104566, Vol. 2*, S.B. Hooker and E.R. Firestone, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, 16 pp.

Vol. 3

McClain, C.R., W.E. Esaias, W. Barnes, B. Guenther, D. Endres, S. Hooker, G. Mitchell, and R. Barnes, 1992: Calibration and Validation Plan for SeaWiFS. *NASA Tech. Memo. 104566, Vol. 3*, S.B. Hooker and E.R. Firestone, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, 41 pp.

Vol. 4

McClain, C.R., E. Yeh, and G. Fu, 1992: An Analysis of GAC Sampling Algorithms: A Case Study. *NASA Tech. Memo. 104566, Vol. 4*, S.B. Hooker and E.R. Firestone, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, 22 pp., plus color plates.

Vol. 5

Mueller, J.L., and R.W. Austin, 1992: Ocean Optics Protocols for SeaWiFS Validation. *NASA Tech. Memo. 104566, Vol. 5*, S.B. Hooker and E.R. Firestone, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, 43 pp.

Vol. 6

Firestone, E.R., and S.B. Hooker, 1992: SeaWiFS Technical Report Series Summary Index: Volumes 1–5. *NASA Tech. Memo. 104566, Vol. 6*, S.B. Hooker and E.R. Firestone, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, 9 pp.

Vol. 7

Darzi, M., 1992: Cloud Screening for Polar Orbiting Visible and IR Satellite Sensors. *NASA Tech. Memo. 104566, Vol. 7*, S.B. Hooker and E.R. Firestone, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, 7 pp.

Vol. 8

Hooker, S.B., W.E. Esaias, and L.A. Rexrode, 1993: Proceedings of the First SeaWiFS Science Team Meeting. *NASA Tech. Memo. 104566, Vol. 8*, S.B. Hooker and E.R. Firestone, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, 61 pp.

Vol. 9

Gregg, W.W., F.C. Chen, A.L. Mezaache, J.D. Chen, J.A. Whiting, 1993: The Simulated SeaWiFS Data Set, Version 1. *NASA Tech. Memo. 104566, Vol. 9*, S.B. Hooker and E.R. Firestone, and A.W. Indest, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, 17 pp.

Vol. 10

Woodward, R.H., R.A. Barnes, C.R. McClain, W.E. Esaias, W.L. Barnes, and A.T. Mecherikunnel, 1993: Modeling of the SeaWiFS Solar and Lunar Observations. *NASA Tech. Memo. 104566, Vol. 10*, S.B. Hooker and E.R. Firestone, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, 26 pp.

Vol. 11

Patt, F.S., C.M. Hoisington, W.W. Gregg, and P.L. Coronado, 1993: Analysis of Selected Orbit Propagation Models for the SeaWiFS Mission. *NASA Tech. Memo. 104566, Vol. 11*, S.B. Hooker, E.R. Firestone, and A.W. Indest, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, 16 pp.

Vol. 12

Firestone, E.R., and S.B. Hooker, 1993: SeaWiFS Technical Report Series Summary Index: Volumes 1–11. *NASA Tech. Memo. 104566, Vol. 12*, S.B. Hooker and E.R. Firestone, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, 28 pp.

Vol. 13

McClain, C.R., K.R. Arrigo, J. Comiso, R. Fraser, M. Darzi, J.K. Firestone, B. Schieber, E-n. Yeh, and C.W. Sullivan, 1994: Case Studies for SeaWiFS Calibration and Validation, Part 1. *NASA Tech. Memo. 104566, Vol. 13,* S.B. Hooker and E.R. Firestone, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, 52 pp., plus color plates.

Vol. 14

Mueller, J.L., 1993: The First SeaWiFS Intercalibration Round-Robin Experiment, SIRREX-1, July 1992. *NASA Tech. Memo. 104566, Vol. 14,* S.B. Hooker and E.R. Firestone, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, 60 pp.

Vol. 15

Gregg, W.W., F.S. Patt, R.H. Woodward, 1994: The Simulated SeaWiFS Data Set, Version 2. *NASA Tech. Memo. 104566, Vol. 15,* S.B. Hooker and E.R. Firestone, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, 42 pp., plus color plates.

Vol. 16

Mueller, J.L., B.C. Johnson, C.L. Cromer, J.W. Cooper, J.T. McLean, S.B. Hooker, and T.L. Westphal, 1994: The Second SeaWiFS Intercalibration Round-Robin Experiment, SIRREX-2, June 1993. *NASA Tech. Memo. 104566, Vol. 16,* S.B. Hooker and E.R. Firestone, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, 121 pp.

Vol. 17

Abbott, M.R., O.B. Brown, H.R. Gordon, K.L. Carder, R.E. Evans, F.E. Muller-Karger, and W.E. Esaias, 1994: Ocean color in the 21st century: a strategy for a 20-year time series. *NASA Tech. Memo. 104566, Vol. 17,* S.B. Hooker and E.R. Firestone, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, 20 pp.

Vol. 18

Firestone, E.R., and S.B. Hooker, 1994: SeaWiFS Technical Report Series Summary Index: Volumes 1–17. *NASA Tech. Memo. 104566, Vol. 18,* S.B. Hooker and E.R. Firestone, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, (in press).

Vol. 19

McClain, C.R., R.S. Fraser, J.T. McLean, M. Darzi, J.K. Firestone, F.S. Patt, B.D. Schieber, R.H. Woodward, E-n. Yeh, S. Mattoo, S.F. Biggar, P.N. Slater, K.J. Thome, A.W. Holmes, R.A. Barnes, and K.J. Voss, 1994: Case Studies for SeaWiFS Calibration and Validation, Part 2. *NASA Tech. Memo. 104566, Vol. 19,* S.B. Hooker, E.R. Firestone, and J.G. Acker, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, (in press).

Vol. 20

Hooker, S.B., C.R. McClain, J.K. Firestone, T.L. Westphal, E-n. Yeh, and Y. Ge, 1994: The SeaWiFS Bio-Optical Archive and Storage System (SeaBASS), Part 1. *NASA Tech. Memo. 104566, Vol. 20,* S.B. Hooker and E.R. Firestone, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, 40 pp.

REPORT DOCUMENTATION PAGE

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing ir structions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden e stimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operatic ns and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washingto 1, DC 20503.

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE July 1994	3. REPORT TYPE AND D/.TES COVERED Technical Memo andum

4. TITLE AND SUBTITLE
SeaWiFS Technical Report Series
Volume 20—The SeaWiFS Bio-Optical Archive and Storage System (SeaBASS), Part 1

5. FUNDING NUMBERS
Code 970.2

6. AUTHOR(S)
Stanford B. Hooker, Charles R. McClain, James K. Firestone, Todd L. Westphal, Eueng-nan Yeh and Yuntao Ge
Series Editors: Stanford B. Hooker and Elaine R. Firestone

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Laboratory for Hydrospheric Processes
Goddard Space Flight Center
Greenbelt, Maryland 20771

8. PERFORMING ORGANIZATION REPORT NUMBER
94B00104

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
National Aeronautics and Space Administration
Washington, D.C. 20546-0001

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
TM-104566, Vol. 20

11. SUPPLEMENTARY NOTES
James K. Firestone, Todd L. Westphal, Eueng-nan Yeh and Elaine R. Firestone: General Sciences Corporation, Laurel, Maryland; and Yuntao Ge: University Space Research Association, Columbia, Maryla id.

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Unclassified–Unlimited
Subject Category 48
Report is available from the NASA Center for AeroSpace Information, 800 Elkridge Landing Road, Linthicum Heights, MD 21090; (301) 621-0390.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

This document provides an overview of the Sea-viewing Wide Field-of-view Sensor (SeaWiFS) Bio-Optical Archive and Storage System (SeaBASS), which will serve as a repository for numerous data sets of interest to the SeaWiFS Science Team and other approved investigators in the oceanographic community. The data collected will be those data sets suitable for the development and evaluation of bio-optical algorithms which include results from SeaWiFS Intercalibration Round-Robin Experiments (SIRREXs), prelaunch characterization of the SeaWiFS instrument by its manufacturer—Hughes/Santa Barbara Research Center (SBRC), Marine Optical Characterization Experiment (MOCE) cruises, Marine Optical Buoy (MOBY) deployments and refurbishments, and field studies of other scientists outside of NASA. The primary goal of the data system is to provide a simple mechanism for querying the available archive and requesting specific items, while assuring that the data is made available only to authorized users. The design, construction, and maintenance of SeaBASS is the responsibility of the SeaWiFS Calibration and Validation Team (CVT). This report is concerned with documenting the execution of this task by the CVT and consists of a series of chapters detailing the various data sets involved. The topics presented are as follows: 1) overview of the SeaBASS file architecture, 2) the bio-optical data system, 3) the historical pigment database, 4) the SIRREX database, and 5) the SBRC database.

14. SUBJECT TERMS
SeaWiFS, Oceanography, Calibration, Validation, Data Set, SeaBASS, SIRREX, Historical Pigment, Bio-optical Data

15. NUMBER OF PAGES
40

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

National Aeronautics and
Space Administration

Goddard Space Flight Center
Greenbelt, Maryland 20771

NASA

POSTMASTER: If Undeliverable (Section 158,
Postal Manual) Do Not Return